Improvement in the γ-ray energy reconstruction of MAGIC and impact on the spectral analysis of the first Gamma Ray Burst detected at TeV energies

Kazuma Ishio

München 2020

Improvement in the γ-ray energy reconstruction of MAGIC and impact on the spectral analysis of the first Gamma Ray Burst detected at TeV energies

Kazuma Ishio

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Kazuma Ishio
aus Osaka, Japan

München, den 18.09.2020

Erstgutachter: Dr. Prof. Thomas Kuhr
Zweitgutachter: Dr. Prof. Masahiro Teshima
Tag der mündlichen Prüfung: 20.11.2020

Contents

Zusammenfassung

Diese Arbeit behandelt die Entwicklung einer neuartigen Methodik (Random Forest basierte Energierekonstruktion, RF-Erec) zur Bestimmung der Energie von sehr hochenergetischen (VHE, Energie $\gtrsim$ 50 GeV) γ-Strahlen-Ereignissen, welche mit den MAGIC-Teleskopen aufgezeichnet wurden. RF-Erec verbessert die Energierekonstruktion von γ-Strahlen und erweitert damit die Möglichkeiten der MAGIC-Teleskope im Vergleich zur früheren Methodik der Energierekonstruktion, die auf Lookup-Tabellen (LUTs-Erec) basiert und in den letzten zehn Jahren verwendet wurde. Wenn die Energierekonstruktion in Bezug auf die Energieauflösung bewertet wird, d.h. der Breite der Gaußschen Anpassung an die Verteilung des mit der Energie normalisierten Schätzfehlers, ist RF-Erec in einem sehr breiten Energie- und Beobachtungszenitwinkelbereich (Zd) um einen Faktor $\gtrsim$ 2 besser, und diese Verbesserung ist bei Beobachtungen unter hohen Zd sogar noch größer. Außerdem ist die Standardabweichung der Fehlverteilung wesentlich kleiner, da der lange Ausläufer, der in der LUTs-Erec zu sehen sind, in der RF-Erec verschwindet. Dies bedeutet, dass die Energiemigrationsmatrix schmaler wird und die Abschätzung der Energie jedes Ereignisses robuster wird. Folglich ermöglicht RF-Erec eine zuverlässige Messung des Spektrums auch in Situationen mit niedriger Statistik, eine ereignisbezogene Analyse wie für Untersuchungen zur Verletzungen der Lorentzinvarianz (LIV) und eine Suche nach spektralen Anomalien. Der Vorteil ist nicht nur eine verbesserte Genauigkeit, sondern auch eine breitere Anwendbarkeit, wie z.B. für Beobachtungen unter hohen Zenitwinkeln und gemeinsame morphologische und spektrale Untersuchungen. Als Nebenprodukt meiner Forschung identifizierte ich auch die Hauptquelle systematischer Unsicherheiten in der LUTs-Erec, legte ihren Mechanismus dar und bestätigte, dass dieser in der RF-Erec unbedeutend ist.

Ich bewertete die tatsächliche Leistungsverbesserung bei der Rekonstruktion der Spektren für verschiedene realistische Szenarien. Einer der Fälle mit der größten Verbesserung ist die Beobachtung einer γ-Strahlenquelle mit sehr steilem Spektrum unter hohen Zd-Winkeln.

Für ein solches Spektrum erschwert die falsche Energierekonstruktion, genauer gesagt das Verschieben von Ereignissen in Bins höherer Energie, die Spektralanalyse erheblich und verringert ihre Zuverlässigkeit. Während sich die Anzahl falsch zu richtig rekonstruierten Ereignissen bei der RF-Erec höchstens auf einen Faktor von einigen wenigen erstreckt, dehnt er sich bei der LUTs-Erec auf mehr als eine Größenordnung aus. Daher werden die mit LUTs-Erec abgeschätzten hochenergetischen Ereignisse von Überlaufereignissen dominiert, wohingegen die RF-Erec den Anteil der echten Ereignisse bei mehr als der Hälfte

hält. Ich zeige, dass die RF-Erec bessere Fähigkeiten als die LUTs-Erec bei der Abschätzung der spektralen Steigung und Amplitude hat sowie zuverlässigere und einheitlichere Ergebnisse als andere verfügbare Strategien für die Spektralanalyse liefert.

Ich habe diese neuartige Methodik in das Standard-Softwarepaket zur MAGIC-Analyse (MARS) implementiert. Es steht nun der MAGIC-Kollaboration zur Verfügung und wird seit dem Jahr 2020 als Teil des Standard-Datenanalysesystems betrachtet.

Die erste wissenschaftliche Anwendung der neuartigen Energieschätzung geschah auf die Daten der MAGIC-Beobachtungen des Gammastrahlenausbruchs (GRB) GRB 180114C. Dies war, nach 15 Jahren intensiver Suche mit den MAGIC-Teleskopen, der erste GRB, der bei VHE γ-Strahlen signifikant nachgewiesen wurde. Das Spektrum hat die steilste Steigung über eine Energiedekade (von 0,2 TeV bis 2 TeV), die jemals mit MAGIC oder einem anderen VHE γ-Strahleninstrument gemessen wurde. Das steile Spektrum ist auf die Absorption durch das extragalaktische Hintergrundlicht (EBL) zurückzuführen, das den Strahlenfluss von γ-Strahlen bei den höchsten Energien auf weniger als ein Hundertstel reduziert. Außerdem wurde die Beobachtung bei hohem Zd durchgeführt. Unter diesen Beobachtungsbedingungen wäre die bisherige Methode zur Energierekonstruktion, LUTs-Erec, nicht in der Lage gewesen, eine zuverlässige Charakterisierung der Form des VHE γ-Strahlenspektrums von GRB 190114C zu liefern. Basierend auf meiner neuartigen Methodik wurden die MAGIC GRB-Daten jedoch erfolgreich analysiert, was zu zwei Nature-Veröffentlichungen führte, die über diese historische Entdeckung berichten (Nature, Bd. 575, p455-458 und p459-463). Die ergiebige Photonenstatistik ermöglicht die Charakterisierung der VHE-Spektren auf kurzen Zeitskalen von nur einer Minute. Die Analyse offenbarte die Existenz einer neuen Emissionskomponente, die sich bis etwa 2 TeV erstreckt. Diese neue Komponente könnte als Synchrotron-Selbst-Compton-Prozess (SSC) aus dem externen Vorwärtsschock des GRB-Ausstroms erklärt werden, der von mehreren Theoretikern seit langem vorhergesagt wurde. Die Daten offenbaren, dass die SSC-Komponente ungefähr das gleiche zeitliche Potenzverhalten wie die Synchrotron-Komponente aufweist, welche mit der Verlangsamung des Schocks abnimmt, und dass sie für einen beträchtlichen Anteil der kinetischen Energie verantwortlich ist, die im Ausstrom des GRBs deponiert wird.

Trotz der technischen Schwierigkeiten beim Nachweis von TeV γ-Strahlen aus GRBs deuten diese Ergebnisse darauf hin, dass die SSC-Emission ein bei GRBs üblicher Prozess sein könnte, was die Notwendigkeit impliziert, unser Wissen über diese extremen Phänomene grundlegend zu aktualisieren.

Abstract

This thesis is about the development of a novel methodology, the Random Forest based Energy reconstruction (RF-Erec), to determine the energy of the very high energy (VHE, energy $\gtrsim$ 50 GeV) γ-ray events detected with the MAGIC telescopes. RF-Erec improves the energy reconstruction of γ-rays, and thereby extends the capabilities of the MAGIC telescopes, compared to the previous methodology for energy reconstruction, which is based on Look-Up-Tables (LUTs-Erec), and has been used over the last decade.

When the energy reconstruction is evaluated in the energy resolution, which is the width of Gaussian fit to the distribution of estimation error normalised with energy, RF-Erec is better by a factor $\gtrsim 2$ in a very wide range of the energies and pointing Zenith distances (Zd). Such improvement is even larger for high Zd observations. Moreover, the standard deviation of the error distribution is substantially smaller, as the long tail seen in the LUTs-Erec disappears in the RF-Erec. This means the energy migration matrix becomes tighter, and the energy estimation of each event becomes more robust. Consequently, RF-Erec enables a reliable spectral measurement even in situations with poor statistics, an event-wise analysis like for Lorentz Invariance Violation (LIV) studies, and a search for anomalies in the spectral shape. The benefit is not only a better accuracy, but also a wider applicability, such as for observations at high Zenith distance, and morphological together with spectral studies. As a side-product of my studies, I also identified the major source of systematic uncertainties in the LUTs-Erec, clarified its mechanism, and confirmed that it is insignificant in the RF-Erec.

I evaluated the actual performance improvement in the spectrum reconstruction for different realistic scenarios. One of the cases with the biggest improvement is on a high Zd observation of a γ-ray source with very steep spectrum. In such spectrum, the energy mis-reconstruction error, namely the spillover to higher energies, complicates substantially the spectral analysis and reduces its reliability. While spillover extends to at most factor of a few in the RF-Erec, it extends to more than one order of magnitude in the LUTs-Erec. Therefore the high energy events estimated using LUTs-Erec are dominated by spillover events, but the RF-Erec keeps the fraction of genuine events to be more than half. I show that the RF-Erec has better ability than LUTs-Erec in estimating the slope and amplitude, as well as more reliable and consistent results among the available strategies for spectral analysis.

I have implemented this novel methodology into the standard MAGIC Analysis and Reconstruction Software (MARS). It is now available to the MAGIC collaboration and, starting from year 2020, regarded as part of the standard data analysis framework.

Abstract

The first scientific application of the novel energy estimation was on the data from the MAGIC observations of the gamma-ray burst (GRB) GRB 190114C. It was the first GRB detected significantly at VHE γ-rays, after more than 15 years of intense searches with the MAGIC telescopes. The spectrum has the steepest shape over one decade in energy (from 0.2 TeV to 2 TeV) that has been ever measured with MAGIC, and with any VHE γ-ray instrument to date. The steep spectrum is due to the absorption by the Extragalactic Background Light (EBL), that reduces the γ-ray flux by factors of several hundreds at the highest energies. Moreover the observation was performed at high Zd. Under these observing conditions, the previous method for energy reconstruction, LUTs-Erec, would not have been able to provide a reliable characterization of the VHE γ-ray spectral shape from GRB 190114C. However, based on my novel methodology, the MAGIC GRB data were analyzed successfully, leading to two Nature papers reporting this historical discovery (Nature, vol.575, p455-458 and p459-463). The rich photon statistics enabled the characterization of the VHE spectra on timescales as short as 1 minute. The analysis revealed the existence of a new emission component extending to about 2 TeV. This new component could be explained as SSC from the external forward shock of the GRB outflow, which has been long predicted by several theorists. The data reveal that the SSC component has approximately the same power-law temporal behavior as the synchrotron component that decreases as the shock decelerates, and that it accounts for substantial amount of the kinetic energy deposited in the outflow from the GRB.

Despite the technical difficulties in detecting TeV γ-rays from GRBs, these results indicate that the SSC emission may be a common process among GRBs, which implies the need to substantially update our knowledge about these extreme phenomena.

Chapter 1

Introduction

The sky is bright and full of stellar objects (stars and galaxies) in the optical light. However it is also bright in other wavebands, including the high energy band, namely the γ-rays. The γ-ray astronomy has become one of the major fields for astronomy in the new era of multi-messenger astronomy. The study of γ-ray sources above $\sim$30 MeV have seen a great progress since the launch of the LAT telescope onboard the *Fermi* satellite in 2008. In the mean time, the energy above $\sim$50 GeV, called very high energy (VHE) γ-ray, has been led by ground-based telescopes, especially by IACTs.

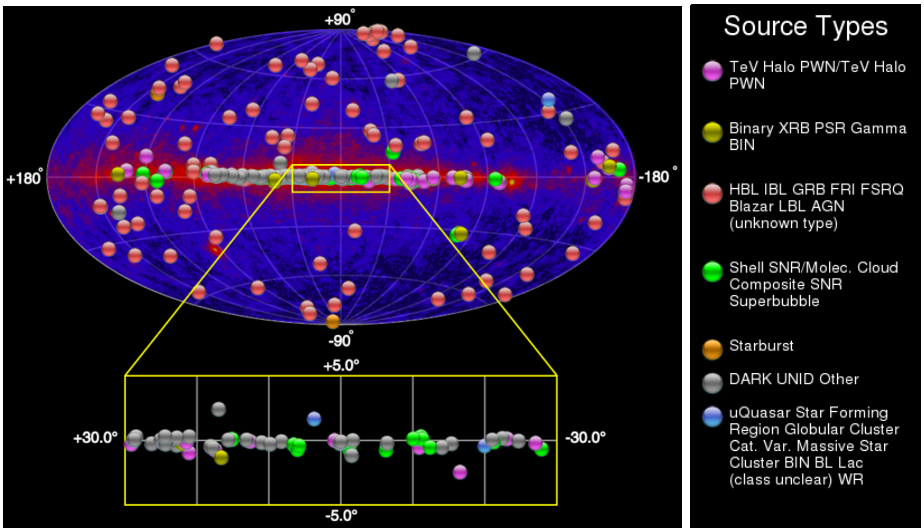


Figure 1.1: Skymap of VHE γ-ray sources

The locations of the VHE γ-ray sources are marked with colored dots on the skymap in a galactic coordinates expressed in Hammer-Aitoff projection. The background is a sky map of the γ-ray flux (low to high is shown as blue to red) measured by the LAT detector onboard *Fermi* satellite. The colors of dots indicate different source types as shown in the legend. From top to bottom, they are, pulsar wind nebulae, binary objects, active galactic nuclei, super nova remnants, starburst galaxies, unidentified objects, and the others. Image taken from http://tevcat.uchicago.edu/

Figure 1.1 shows the map of VHE γ-ray sources, superposed on the HE (GeV) γ-ray skymap from *Fermi*. There have been more than 200 sources discovered at VHE γ-rays, which comprise various kinds of objects for a wide range of scales and distances, all of which are energetic phenomena. The galactic objects, which are mostly along the galactic plane of our galaxy, are related to the explosive events of stars like supernova remnant (SNR) and pulsar wind nebulae (PWN), and the extra-galactic objects, which scatters throughout the map, are mainly super massive objects as active galactic nuclei (AGN).

All of them must contain extreme environments to produce such high energy photons. They must be extraordinary cosmic accelerators. Then the interest is how the γ-rays are radiated. A key ingredient to obtain is the SED, namely the power distribution of the emission as a function of the photon's energy. Comparing the broad-band spectra together with other wavelengths measured by various telescopes, we can clearly see that the emission from those VHE γ-ray sources are of non-thermal origin, which requires the existence of particles in extremely high energy. In some cases, the particles are accelerated to energies millions of times higher than those that can be produced in Earth-based laboratories.

Figure 1.2 shows two examples of such non-thermal emissions. The first spectrum is from an AGN named Markarian 501 (Mrk 501), and the second from a SNR named Cassiopeia-A (CasA). By investigating the γ-ray spectra we can unravel the extreme processes that occur in these cosmic sources. More specifically, the spectral shapes, e.g. spectral slope, spectral break, and cut-off, bring information about the emission mechanisms, which relate to different parent particle populations and their environments. Typically, the mechanisms are identified to be leptonic interactions, in which VHE γ-rays are produced through inverse Compton boosting of ambient photons, or to be hadronic interactions, in which the decay of neutral pions produced VHE γ-rays. In the two examples mentioned above, the γ-ray emission has been identified to have a leptonic origin (electrons play a major role) for Mrk501, and a hadronic origin (protons play a major role) for CasA. Moreover, the γ-ray spectral shape also reflects the parent particles' energy distribution, the magnetic field, the ambient medium, and the speed of the accelerator, namely the shock wave. And in some cases, the γ-ray spectra have a temporal evolution or spatial morphology, which provide additional dimensions to be explored. Therefore the energy distribution of γ-rays is a crucial tool to understand the non-thermal emission processes occurring in the extraordinary cosmic accelerators.

We can also study the environment where the radiation is propagated, through the interactions of the γ-rays with the ambient medium. The most important example is the measurement of the diffuse extragalactic background light through the energy-dependent absorption of the γ-rays that travel to Earth from sources located at cosmological distances. Additionally, the γ-rays can also be used to perform a wide range of fundamental physics investigations. Relevant examples of these applications are the searches for new particles beyond the Standard Model, e.g., the elusive Dark Matter (DM) and the Axion [109, 115]. As shown in figure 1.3, the energy distribution plays an essential role for these studies. Moreover, γ-rays can also be used to perform tests of Lorentz Invariance Violation (LIV), which depend on the energy and the arrival time of the detected photons [14, 83].

The work presented in this thesis is dedicated to the improvement of the en-

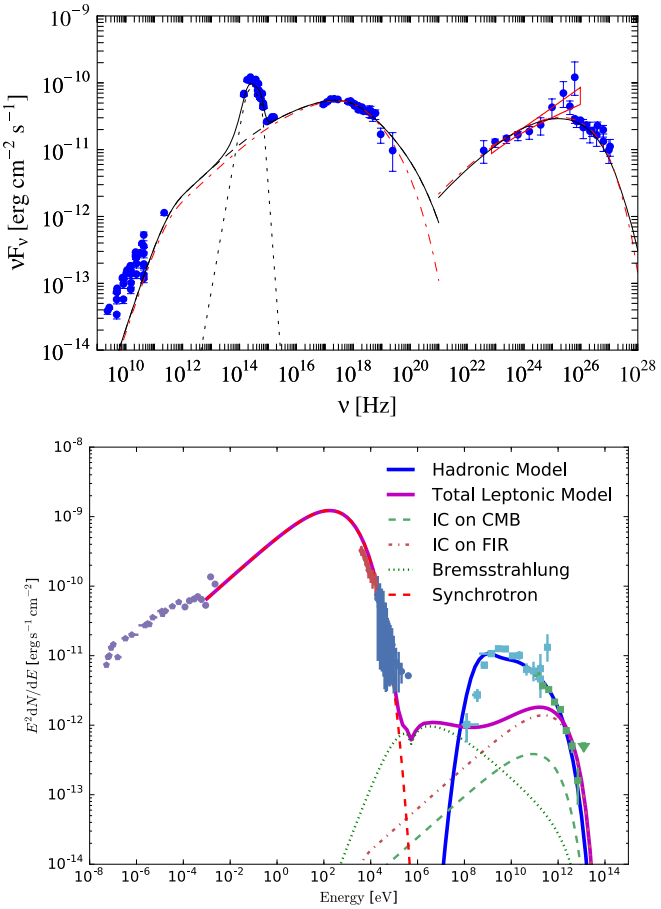


Figure 1.2: Broadband spectrum dominated by non-thermal emission from γ-ray sources

Top: The broadband emission spectrum of Mrk 501, taken simultaneously by multiple telescopes. The red bow-tie in the figure corresponds to the 68% containment of the power-law fit to the average *Fermi*-LAT spectrum (photon index 1.74 ± 0.05). The dotted black curve denotes the fit to the starlight emission of the host galaxy assuming a template of a luminous elliptical galaxy. The black curves and red dot-dashed curves are the two example model spectra with different model parameters. The model is leptonic and both consist of two emission components, synchrotron and synchrotron-self-compton (SSC). Figure extracted from [5].
Bottom: The broadband emission spectrum of CasA. The different lines show the results of fitting the measured energy fluxes using naima (https://github.com/zblz/naima) and assuming a leptonic or a hadronic origin of the GeV and TeV emission. The hadronic origin was successfully confirmed by the fit to the spectral data points, which preferred exponential cutoff over power law at the highest energy. Figure extracted from [12].

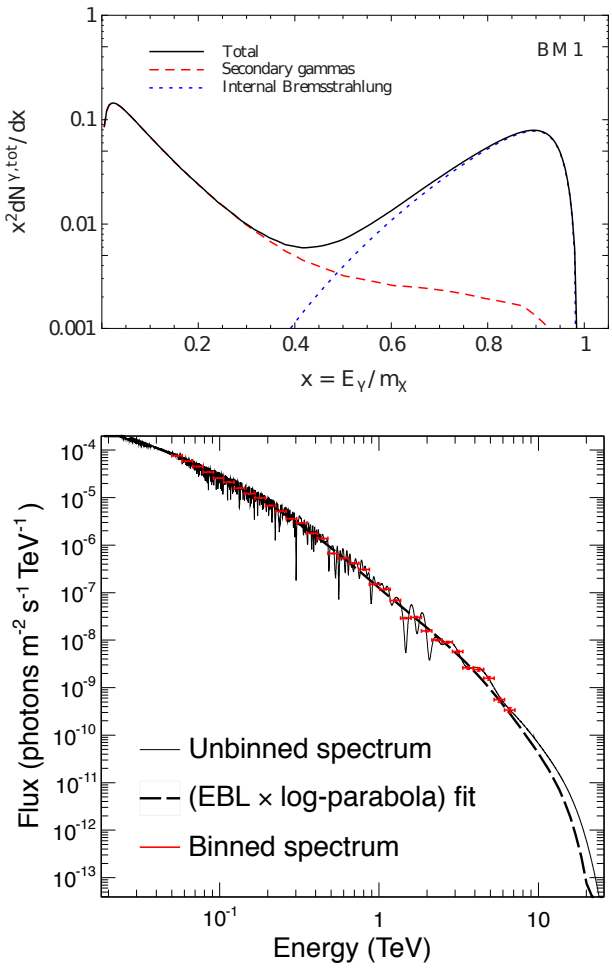


Figure 1.3: The expected effect by Dark Matter and Axion on the observed spectrum of γ-ray

Top: Simulation of the γ-ray spectrum for a benchmark model of γ-ray emission from Dark Matter (DM). The model represents a typical region in the supersymmetric parameter space where internal-bremsstrahlung effect becomes important. The horizontal axis is the γ-ray energy in the unit of the assumed mass of DM m_χ, where $m_\chi \sim 1.4$ TeV. Figure extracted from [37].
Bottom: Simulation of the observation of one γ-ray source at $z = 0.1$, with the effect of γ-ALP mixing. Figure extracted from [115].

ergy estimation of the γ-rays in the data from the Major Atmospheric Gamma-ray Imaging Cherenkov telescope system (MAGIC). MAGIC is a worldwide leading instrument for γ-ray astronomy in the energy range from $\sim$ 25 GeV to beyond 100 TeV. It is a system of two IACTs with a 17 m diameter parabolic reflector. The two telescopes are used to record stereoscopic images with the Cherenkov radiation from the relativistic secondary charged particles from the extensive air showers (EAS) induced by the VHE γ-rays and cosmic rays penetrating the Earth atmosphere. From the characteristics of the images, it is possible to estimate the properties of the primary particle from each EAS event, such as the particle species (γ-ray or cosmic ray), the incoming direction and the energy.

Chapter 2

Detection of VHE γ-rays

The atmosphere is opaque to VHE γ-rays and prevents them to arrive at ground level. However, the detection from ground is possible by measuring a cascade of numerous particles produced during the absorption. The cascade is called "Extensive Air Shower (EAS)" and it is common process for a high-energy particle to impinge the atmosphere. Among various techniques to measure the EAS, this study focuses on the "Imaging Atmospheric Cherenkov Telescope (IACT)", which is the most successful technique to detect VHE γ-rays. The IACT technique takes advantage of the Cherenkov radiation from the cascaded particles in EAS, which makes the shape of EAS visible.

This chapter introduces the detection method of VHE γ-rays in the following steps. First of all, since all the processes are governed by the atmosphere, the related characteristics of atmosphere are introduced in section 2.1. Section 2.2 explains the properties of EAS, and the radiation from EAS will be followed by section 2.3. In the last section, section 2.4, the IACT technique will be described.

2.1 Atmosphere

Since the Extensive Air Shower (EAS) occurs in the atmosphere, it is important to understand the atmospheric properties that affect the development of the shower. In addition, the attenuation and scattering of light in the atmosphere are also important to IACT, which measures the radiation from EAS.

2.1.1 Density profile

Although the actual atmosphere is a complex turbulent system, it is still adequate to simplify the profile for the development of EAS. The atmosphere extends up to the altitude of 100 km[1] and composed of the four primary layers, which are, from bottom on the ground to top, the troposphere, stratosphere, mesosphere, and thermosphere. The lowest part, troposphere, is of greatest importance to EAS development, as well as to the radiation and the propagation of Cherenkov light, because three quarters of the atmosphere's mass resides within

[1] This is named as Karman line and is commonly used to define the boundary between Earth's atmosphere and outer space.

the troposphere. The extension is from sea level to around 11 km height, and the density profile follows a exponential function,

$$\rho(h) = \rho_s \exp(-h/H_0) \tag{2.1.1}$$

where H_0 is called scale height and ρ_s is the density at sea level. However the density profile varies mainly depending on the locations and the seasons. Figure 2.1 shows how well the density profile in case of La Palma follows the exponential law [56]. The number density $n(h)$ instead of the density $\rho(h)$ is measured and the vertical axis is the ratio to the exponential profile $n_s \exp(h/H_0)$, where $n_s = 2.55 \cdot 10^{19}$ cm^{-3} is the U.S. Standard Atmosphere density and $H_0 = 9.5$ km is an average scale height. The variation of the atmosphere profile causes a systematic effect in the measurement of γ-rays.

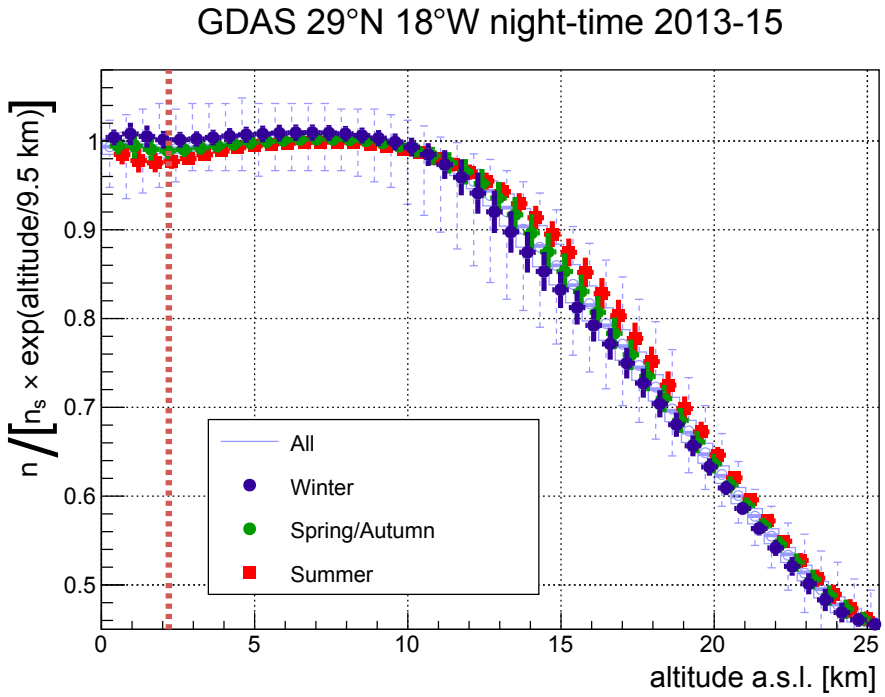


Figure 2.1: Seasonal change of atmospheric density profile

Density profiles of 3-years (2013 - 2015) Global Data Assimilation System[2] nowcasts at 21h, 0h, 3h and 6h UTC. The dark blue, red and green points display an "average Winter", "average Spring/Autumn" and "average Summer" scenario, respectively[3]. Error bars represent the standard deviation, while the light blue lines show the peak-to-peak differences for the entire data set. Figure extracted from [56] and modified.

2.1.2 Slant depth

When a high-energy particle is absorbed in matter, the absorption process progresses as the amount of matter it passes. The path of the particle varies

dramatically in the density from the almost absolute emptiness of extragalactic space to the relative clutter of our atmosphere, to the extreme density of our rocky Earth or even lead shielding. Therefore, measure of the path length is expressed as the atmospheric depth along a line inclined to the vertical. This is called the "atmospheric slant depth". In this thesis, the slant depth is denoted as X. For the inclined paths at the Zenith distance θ_z less than 75°, where the effect of the curvature of the Earth is small, the slant depth as a function of the height h can be approximated to be

$$X(h, \theta_z) = X_s \exp(-h/H_0)/\cos\theta_z \tag{2.1.2}$$

where $X_s = 1033$ g·cm^{-2} is the vertical atmospheric depth at sea level. The standard values of slant depth in vertical path (vertical depth) are shown in table 2.1.

Table 2.1: Atmospheric parameters

altitude (km)	vertical depth (g/cm^2)	Cherenkov threshold (MeV)	Cherenkov angle (″)
40	3	386	0.076
30	11.8	176	0.17
20	55.8	80	0.36
15	123	54	0.54
10	269	37	0.79
5	550	28	1.05
3	715	25	1.17
1.5	862	23	1.26
0.5	974	22	1.33
0	1032	21	1.36

Atmospheric parameters of relevance to Cherenkov light production by electrons. The parameters are given for the US standard atmosphere, measured relative to sea level. Table extracted from [52].

2.1.3 Atmospheric absorption of light

The absorption of light by the atmosphere is due to different attenuation and scattering processes, and their strengths are dependent on the wavelength of the light to absorb. Figure 2.2 shows the transmissions due to individual absorption processes, together with the total transmission.

The wavelength below 242 nm suffers from the disassociation of normal oxygen (O_2). At wavelengths of less than 290 nm the ozone component of the atmosphere strongly absorbs ultraviolet [102]. Wavelengths larger than 800 nm

[2] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas

[3] The seasons indicated here are for the following months; Winter is from January to April, Spring is May and June, Autumn is October and November, Summer is July, August and September.

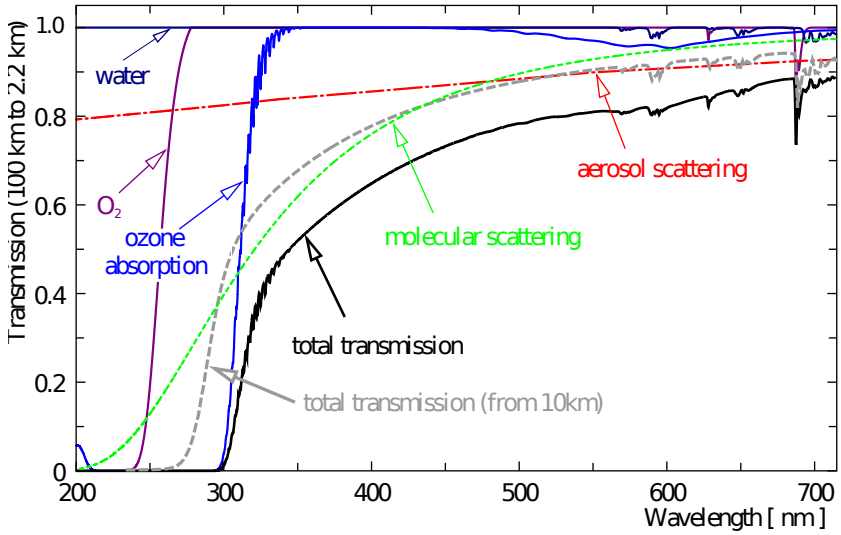


Figure 2.2: The atmospheric transmission

Direct transmission of light from space (100 km altitude) along a vertical path to an altitude of 2.2 km (the same as that of MAGIC telescopes) is calculated with MODTRAN. Together with the total transmission (black), the colored lines indicate the transmissions due to atmospheric absorption by ozone absorption (blue), molecular scattering (Rayleigh scattering, green) and aerosol scattering (Mie scattering, red). The total transmission for the light from 10 km of altitude is also shown as the grey dashed line. Figure extracted from [27] and modified by [50].

(infrared) are usually absorbed by the H_2O and CO_2 molecules in the air. In the region between them, the primary mechanism for attenuation is scattering by the atmospheric molecules themselves (Rayleigh scattering) and by natural and man-made aerosol particles (Mie scattering).

The cross section of Rayleigh scattering is proportional to λ^{-4}, thus the shorter part of the spectrum is especially affected by the scattering. Mie scattering is quite complex. In general it is not possible to calculate the total aerosol extinction coefficient analytically. Moreover, aerosol scattering depends on particle composition, which can change quite rapidly depending on the wind and weather conditions [6]. A much more realistic model of aerosol vertical structure, aerosol properties plus all the relevant molecular absorption and scattering is included in the MODTRAN program [27]. When good visibility (40 km), the contribution of Mie scattering is less than 20% of Rayleigh scattering [102].

2.2 Extensive Air Shower (EAS)

EAS is common phenomenon to all the high-energy particles (cosmic rays or γ-rays)[4] impinging the atmosphere. When a high energy particle (primary par-

[4] High-energy particles with intrinsic mass are known as "cosmic" rays, while photons, which are quanta of electromagnetic radiation (and so have no intrinsic mass) are known by their common names, such as γ-rays or X-rays, depending on their photon energy.

ticle) enters the atmosphere, the interaction of the primary particle with atmospheric matter create new particles (secondary particles). The interaction will repeat also to the secondary particles, as long as the energy of the particle is sufficiently high (the lowest energy called critical energy). In this way a cascade of particles will be formed, and the repetitive interaction creates numerous particles. In addition to the deceleration via ionization, the multiplication of secondary particles reduces individual energies and they are then stopped in the matter and absorbed.

While the EAS from VHE γ-ray is the target to measure in this study, the EAS is common phenomenon to all the high-energy particles impinging the atmosphere. These non-γ-ray showers produce a very high level of background events that mask the rare γ-ray showers that account for $\ll 1$ % of all showers, making it very difficult to filter them out. Therefore it is essential to understand the properties of EAS to discriminate the EAS between different primary particles. The EAS is classified to two distinctive types. One is called electromagnetic shower, whose primary particle is a photon (VHE γ-ray) or a leptonic particle like electron. The other is called hadronic shower, whose primary particle is a hadronic particle like a nucleon or a heavier nucleus, especially proton and helium.

Hadronic showers cause the main background events in the detection of VHE γ-rays. However, hadronic showers have different properties from electromagnetic showers, and therefore it is possible to mostly remove by taking advantages of the differences. In addition, the electromagnetic showers are geometrically simple and thus helps the direction reconstruction and the energy estimation of primary particle.

Following the motivations above, the properties of electromagnetic showers and hadronic showers will be elaborated separately in section 2.2.1 and section 2.2.2.

2.2.1 Electromagnetic Shower

Dominant processes

The dominant interaction processes in an electromagnetic shower are bremsstrahlung and pair production[5]. Bremsstrahlung is radiative loss process for electrons and dominant above $E_c = 87\,\mathrm{MeV}$ in the atmosphere. The loss increases almost in proportion to energy for $E \gg E_c$ (figure 2.3a), which is characterised as

$$\frac{dE_e(X)}{dX} = -\frac{1}{X_0} E_e(X) \tag{2.2.1}$$

where X_0 is called the radiation length, $37\,\mathrm{g/cm^2}$ for electrons in air.

Pair production is the process where a photon vanishes and in its place an electron and a positron appear as a matter-antimatter pair of particles. This is possible in photon-matter collisions for the photon energy above $2m_e = 1.022\,\mathrm{MeV}$, and becomes dominant above $\sim 10\,\mathrm{MeV}$. In the case of high energy limit, the pair-production probability reaches constant value $\sim 7/9$ thus

[5] Note that γ-ray-induced showers might contain muons, although the number of muons in a γ-ray-induced shower is much smaller than that in a hadron-induced shower.

the pair-production cross section per target air nucleus is

$$\sigma_{\gamma\to e^+e^-} \sim \frac{7}{9}\frac{A}{N_A X_0} \tag{2.2.2}$$

where A is the averaged mass number in the air and N_A is Avogadro number. [52] The mean free path $\lambda_{\gamma\to e^+e^-}$ is

$$\lambda_{\gamma\to e^+e^-} = \frac{1}{n\sigma} = \frac{A}{N_A}\cdot\frac{9}{7}\frac{N_A X_0}{A} = \frac{9}{7}X_0 \tag{2.2.3}$$

Therefore in pair creation, the energy transfer of losing roughly half energy occurs after traveling $\frac{9}{7}X_0$, which can be regarded as exponential energy loss at similar rate to bremsstrahlung.

Development of the cascade

In case of a γ-ray entering the atmosphere, the first interaction is predominantly pair production. It occurs typically at altitude around 20 km for a vertical shower, where the mean free path eq.(2.2.3) corresponds to the altitude as shown in table 2.1.

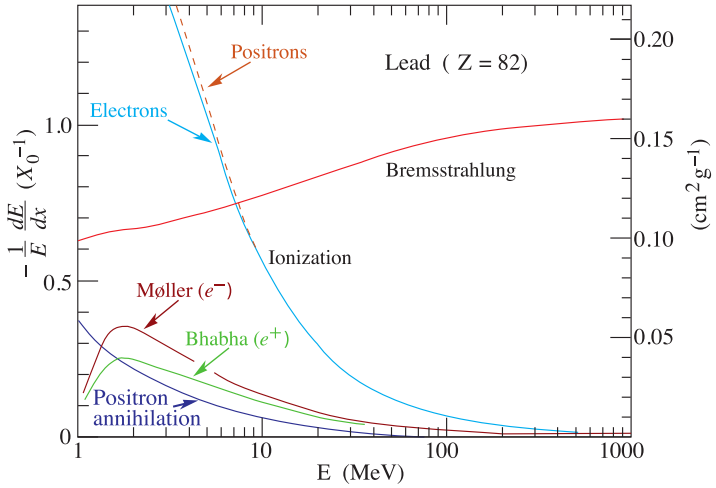
The electron-positron pair then produce additional photons through bremsstrahlung, some of which are still high energy that can produce electron-positron pair. This cycle repeats with the total number of particles in the shower increasing exponentially. However, as the average energy of each particle decreases due to the increase of number of particles, the particles dissipate their energy by ionization and excitation rather than by the generation of more shower particles. Eventually the number of particles will reach the maximum and this point is known as the "shower maximum" (The correspondent vertical height from the ground is called MaxHeight). Afterwards the number of particles decreases and the shower finishes.

Longitudinal distribution

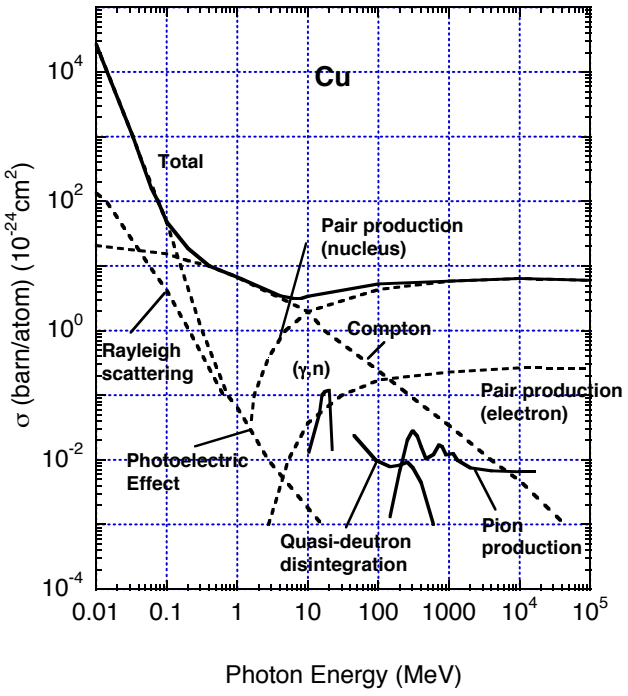
The longitudinal distribution of the secondary particles is one of the most distinctive property to characterise an EAS. Since electromagnetic showers are governed by the two interactions with exponential energy losses at similar rate, the cascade process can be illustrated in a simple binomial energy transfer model called Heitler model.

In the sketch shown in figure 2.4, each line segment can be thought of as a particle or as a packet of energy. At each vertex the energy on a line is split in two. Branching occurs after one collision length λ for either pair production or bremsstrahlung. The number of particles at the depth of shower X develops as $N(X) = 2^{X/\lambda}$. The multiplication of the number of particles continues until they can not produce γ-rays any more at which point the energy falls below the critical energy for bremsstrahlung E_c. When the multiplication reaches the shower maximum $X_{\max}$, the number of particles is $N(X_{\max}) = E_0/E_c$. On the other hand, the shower maximum in this model can also be expressed as a function of the primary energy E_0, and is $X_{\max} = \lambda\frac{\ln(E_0/E_c)}{\ln 2}$. The two basic features we obtain are

$$N_{\max} \propto E_0 \quad \text{and} \quad X_{\max} \propto \ln(E_0) \tag{2.2.4}$$



(a) Electron cross section



(b) Photon cross section

Figure 2.3: Electron and photon cross sections from various processes

(a)Electron cross section and energy loss in Lead, as a function of energy. Figure extracted from [106].

(b) Total photon cross section in Cu, as a function of energy. Figure extracted from [68].

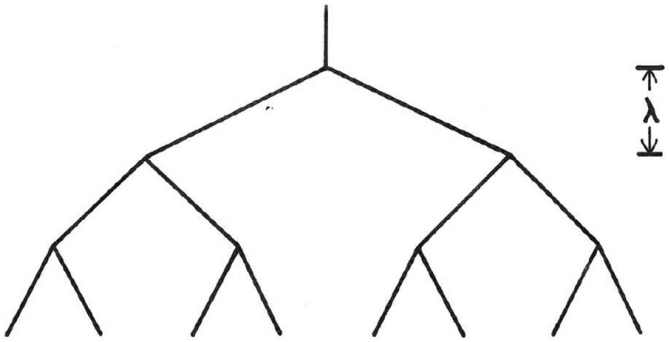


Figure 2.4: A simple model for an electromagnetic shower.

Image taken from [52].

Although the model is extremely simple, these two qualitative features are applicable to the development of an electromagnetic shower, as well as to the electromagnetic cascade parts of a hadronic shower [104].

The expected first interaction point would not be different because the photon cross section to pair creation is almost flat with energy [68]. On the other hand, the simple model for the electromagnetic shower shows the development of shower development dependent on the primary energy, where the higher the energy, the deeper the shower, with the depth roughly following $\log_2(E/E_0)$. Thus the MaxHeight lowers as the energy is higher on average.

There are detailed mathematical treatments of the electromagnetic shower development, as the solution of transport equations taking into account particle loss as well as particle production. The model best described is proposed by Rossi and Greisen [95][6].

The development of the shower from γ-ray with primary energy E_0 can be denoted in terms of the number of electrons and positrons $N_e(t)$ as

$$\begin{aligned} N_e(t) &\sim \frac{0.31}{\sqrt{\ln(E_0/E_c}\exp\left[t(1-\frac{3}{2}\ln s)\right]} \\ s &= \frac{3t}{t+2\ln(E_0/E_c)} \end{aligned} \tag{2.2.5}$$

where s is called the age parameter, the shower maximum occurs for $s = 1$ and t is the depth of shower in the unit of radiation length $X_0 = 37\mathrm{g/cm}^2$, namely, $t = X(\mathrm{g/cm}^2)/X_0$. Examples of the shower development from various initial energy are shown in figure 2.5. The shower extends ~ 10 km longitudinally, and the extension is longer with higher energy. Since the pair creation cross section is constant independent from the energy for $\gtrapprox 1$ GeV, the higher energy shower grows deeper in the atmosphere, while the expected first interaction point stays the same. Therefore MaxHeight (the height of shower maximum) is on average

[6] They propose two approximate solutions, Approximation A and Approximation B. Approximation A neglects the electron energy loss in ionization. Approximation B, which is introduced here, accounts for the ionization loss and thus depicts the development dependent on the primary energy.

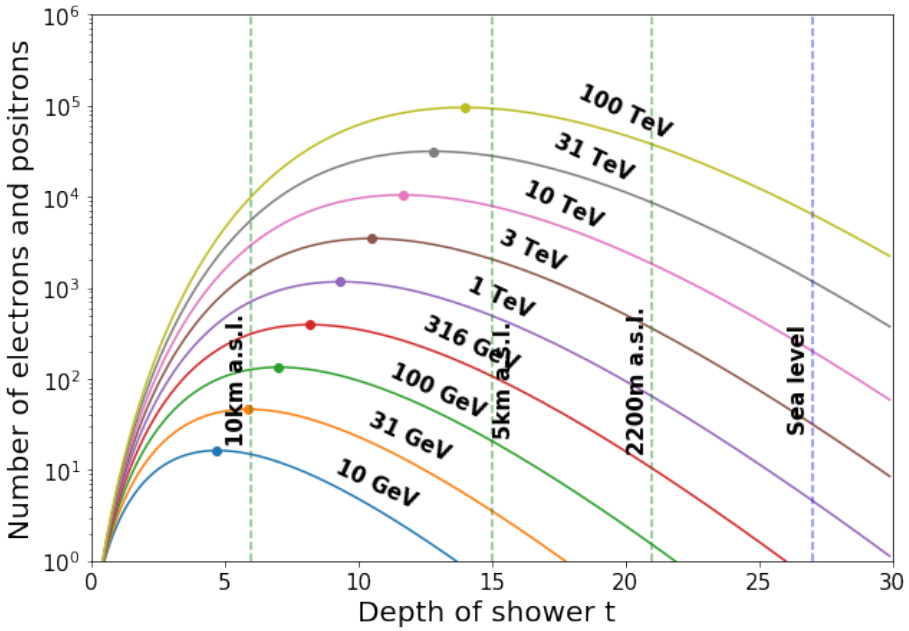


Figure 2.5: Development of photon-initiated showers with different initial energy.

The lines in the different colors depict the development of the showers with various initial energy E_0, with the dots on the lines indicating shower maximum ($s = 1$). The horizontal axis is the depth of the shower t, in the unit of the radiation length X_0. The vertical lines show the depth of the showers at 2200 m a.s.l.(altitude of MAGIC telescopes) and sea level. The shower is expected to start at the same height, as the cross section of the pair creation for the incoming γ-ray is constant regardless of the initial energy.

related to the energy of initial γ-ray. Given that the measurement of the shower maximum is essential to observation of EAS, it is possible to observe the shower with the energy until about 100 TeV when it is observed towards zenith, because the shower maximum is at ~ 5 km above sea level and visible from ground level.

In reality, the longitudinal distribution suffers from fluctuation. The fluctuation of the longitudinal profile in $\ln N_e$ is

$$\Delta(\ln N_e(s)) \sim \frac{\Delta N_e(s)}{N_e(s)} \sim \frac{9}{14}(s - 1 - 3\ln s) \tag{2.2.6}$$

While the fluctuations in shower size are small near the shower maximum, the growing and decaying phase of shower would show large fluctuations.

In addition to the fluctuation in the development in atmosphere, the fluctuation also comes from the first interaction point. The fluctuation of shower profiles suffering from both effects is well depicted when it is compared in several showers simulated by Monte Carlo study, as shown in figure 2.6. The number of secondary particles are very different between the showers at depths smaller or larger than the shower maximum point. However, the number of secondary particles stays similar at their shower maximums even taking into account the fluctuated interaction points.

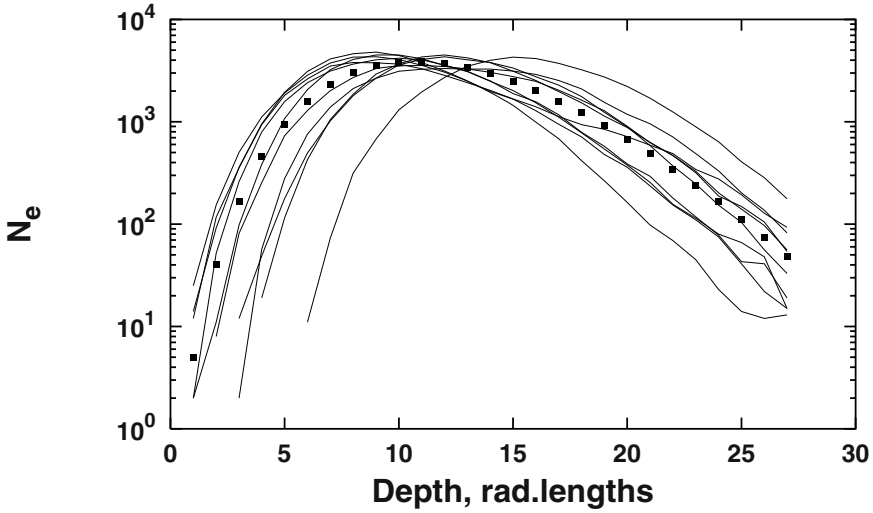


Figure 2.6: Fluctuations in the longitudinal development of EAS

Number of electrons (size) with energy above 1 GeV as a function of the shower depth in the unit of the radiation length X_0, in a shower from primary photon of energy 10^5 GeV. The lines show the profiles of 10 examples of individual showers and the squares show the average profile from 100 showers. At the maximum of the averaged shower profile, all the example profiles except for one take about the same size. On the other hand, the sizes are very different between the showers at depths smaller than 5 and greater than 15 radiation lengths. Figure extracted from [104].

Lateral distribution

The shower develops along the primary γ-ray direction, however the particles created in the shower have lateral distribution not only because of the opening angles in pair creation and bremsstrahlung, but mainly because of multiple Coulomb scattering. The lateral spread of a shower is determined largely by Coulomb scattering of the many low-energy electrons [106] and is characterised by the Molière unit, which is the spread of the particle with the energy E_c in a shower.

$$\begin{aligned} X_M &= \frac{21\text{MeV}}{E_c} X_0 \sim 9.3\text{g/cm}^2 \\ R_M &= X_M/\rho \end{aligned} \tag{2.2.7}$$

Where Molière radius is R_M, converted from Molière unit X_M through the density in the atmosphere ρ. $R_M \sim 80$ m at sea level and $R_M \sim 200$ m at the position of shower maximum (for a vertical 1 TeV shower). In case of an electromagnetic shower, the development has been studied both by an analytical approach to solve the coupled cascade equations by Nishimura and Kamata [90] and by empirical fits to Monte Carlo simulations [10]. The full function $\rho_e(r)$, the density of electrons and positrons as a function of the radial distance r from shower axis, is known as the Nishimura–Kamata–Greisen (NKG) lateral distribution function and is given by [102]

$$\rho_e(r) \sim \frac{N_e(t)}{2\pi R_M^2} \frac{\Gamma(4.5-s)}{\Gamma(s)\Gamma(4.5-2s)} \left(\frac{r}{R_M}\right)^{s-2} \left(1+\frac{r}{R_M}\right)^{s-4.5} \tag{2.2.8}$$

where $\Gamma(x)$ is a gamma function.

In case of 1 TeV photon, the lateral distribution at shower maximum $s = 1$ is figure 2.7. The figure shows electrons are condensed within $\sim$ 10 m of radial distance from shower axis. The shower is highly collimated along the incident direction of the primary particle.

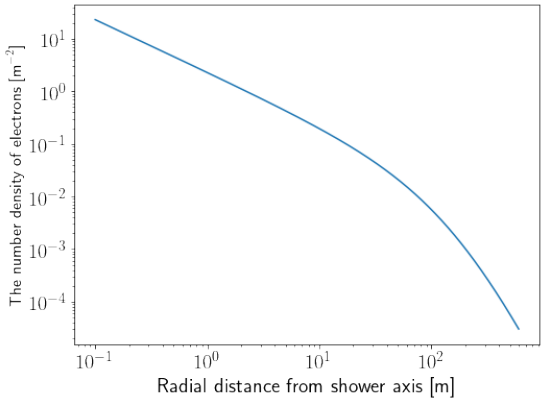


Figure 2.7: Lateral distribution of an electromagnetic shower from 1 TeV photon

The electron density as a function of the radial distance from the shower axis, for an air shower of 1 TeV photon, at the position of shower maximum ($s = 1$, $R_M = 200$ m).

2.2.2 Hadronic Shower

The hadronic shower has the hadronic cascade as the parent process that is chiefly responsible for the energy transport within a shower. The hadronic cascade produces various kinds of primary fragments and mesonic particles, which further decay into different particles. Thus the development of a hadronic shower is quite complex with many kinds of secondary particles and the structure of the cascade largely fluctuates.

In this study, the hadronic shower needs to be understood in the aspect of background events for the VHE γ-ray detection. The development of hadronic shower is focused especially on the difference of morphology compared to the electromagnetic shower. And the other properties related to VHE γ-ray detection are mentioned as well.

Development of the cascade

Figure 2.8 is the sketch of the main processes in the hadronic shower.

When a hadronic high-energy particle enters the Earth's atmosphere, it interacts with a nucleus in the air (mainly nitrogen, oxygen, and argon) via strong

force, at a typical height of 15 to 35 km and initiate the successive collisions producing a shower of secondary particles. The productions are not only baryons like protons and neutrons, but also mesons, great majority of which are charged and neutral pions and partially kaons (K mesons). The neutral pions immediately decay mostly into photons[7]

$$\pi^0 \rightarrow 2\gamma \tag{2.2.9}$$

and the daughter γ-rays start electromagnetic cascades practically at the interaction point, which are called sub-showers. The charged pions decay in the following way.

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{2.2.10}$$
$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \tag{2.2.11}$$

Due to the longer life time compared to neutral pions, the charged pions above the critical energy of $E_c^\pi = 20$ GeV [8] continue as part of the hadronic cascade, contributing to the further interaction [46, 59, 78]. High energy proton continues the hadronic cascade until the energy of the projectiles falls below the one-pion threshold of ~ 280 MeV. The atmosphere has the thickness as a calorimeter for the collisions with a vertical thickness of more than 11 interaction lengths for hadronic cascades and 26 radiation lengths for electromagnetic cascades [78].

The processes involved are much more complicated than an electromagnetic shower, therefore the shower profile is also more fluctuated and complicated, as shown in figure 2.9.

Approximation of the development

The complexity of the hadronic shower process can be simplified in the concept of the superposition of the multiple layers of cascades; the primary hadronic cascade is overlapped by pion cascades and electromagnetic cascades generated from the neutral pions from the hadronic cascades, as shown in figure 2.8.

In the simplified picture, the development of the hadronic skeleton of the shower is carried by hadronic component consisting of various kinds of nucleons and mesons, and the skeleton works as a source of pions, which are the lightest hadron thus dominant production[9]. The pions are traced in the model as they are the sources of electromagnetic sub-showers and muons component. The production rate is $\pi^+ : \pi^- : \pi^0 = 1 : 1 : 1$, and the model assumes that the pions are produced after each interaction length by $2N_\pi$ of charged pions and N_π of neutral pions, where $3N_\pi$ is pion multiplicity. For pion energy between 1 GeV and 10 TeV a charged multiplicity $2N_\pi \sim 10$.

The charged pions continue as part of the hadronic cascade, contributing to the further interaction before their decay, as long as their energy is above the

[7] Neutral pions decay immediately into two γ-rays unless their energy is extremely high. The neutral pion decay length is $l_d = \gamma_{\pi^0} \times 2.51 \times 10^{-6}$ cm.

[8] Charged pions have a much longer decay length $l_d = \gamma_{\pi^\pm} \times 780$ cm - they can either decay or re-interact. This interaction/decay competition of all charged mesons determines the details of the development of hadronic showers. High energy charged pions, with their large decay length because of time dilation, almost exclusively interact. Low energy pions decay into muons and muon neutrinos.

[9] The second dominant production is kaons, but the ratio of kaons to pions is an asymptotic value of ~ 0.1 [59].

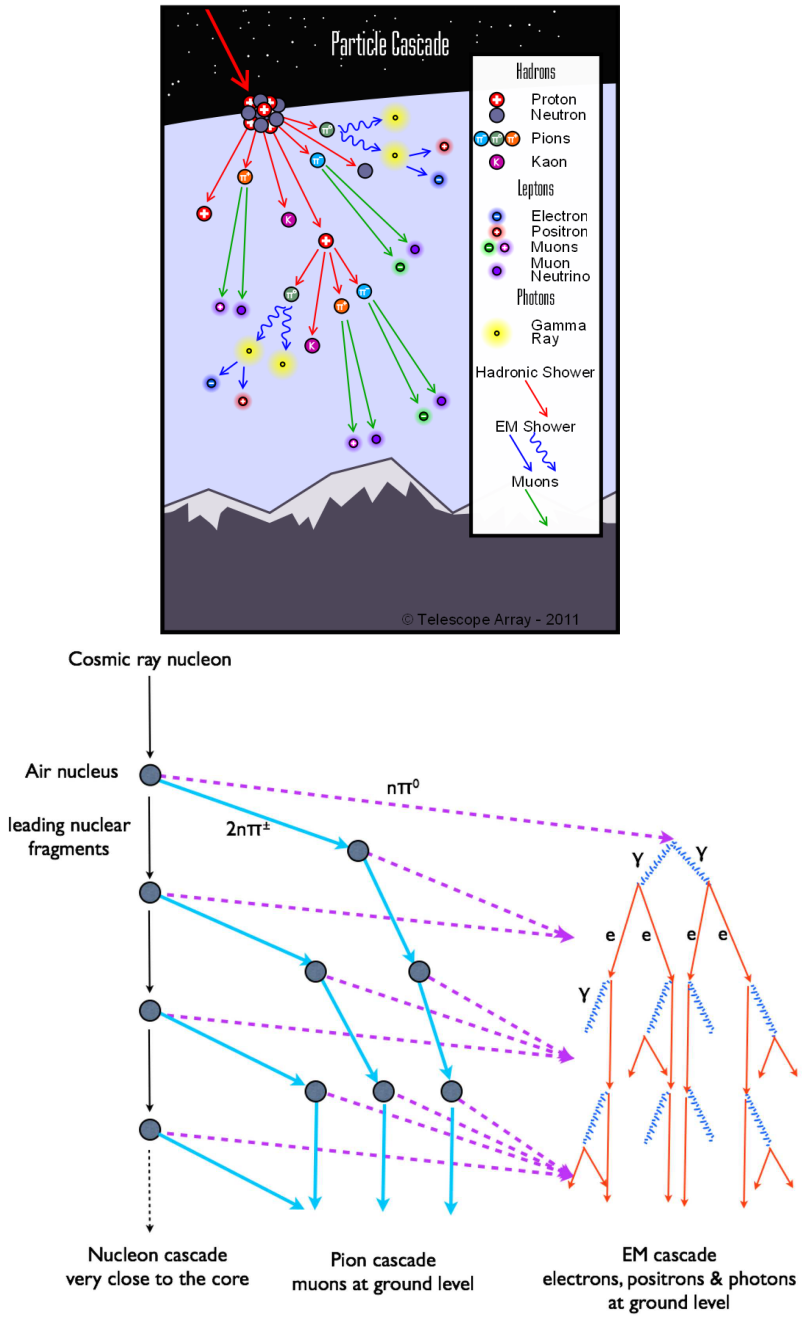


Figure 2.8: Development of a hadronic shower in the atmosphere

A sketch of hadronic interactions (top, image taken from http://www.telescopearray.org) and simplified model of hadronic air shower (bottom, Figure extracted from [78]).

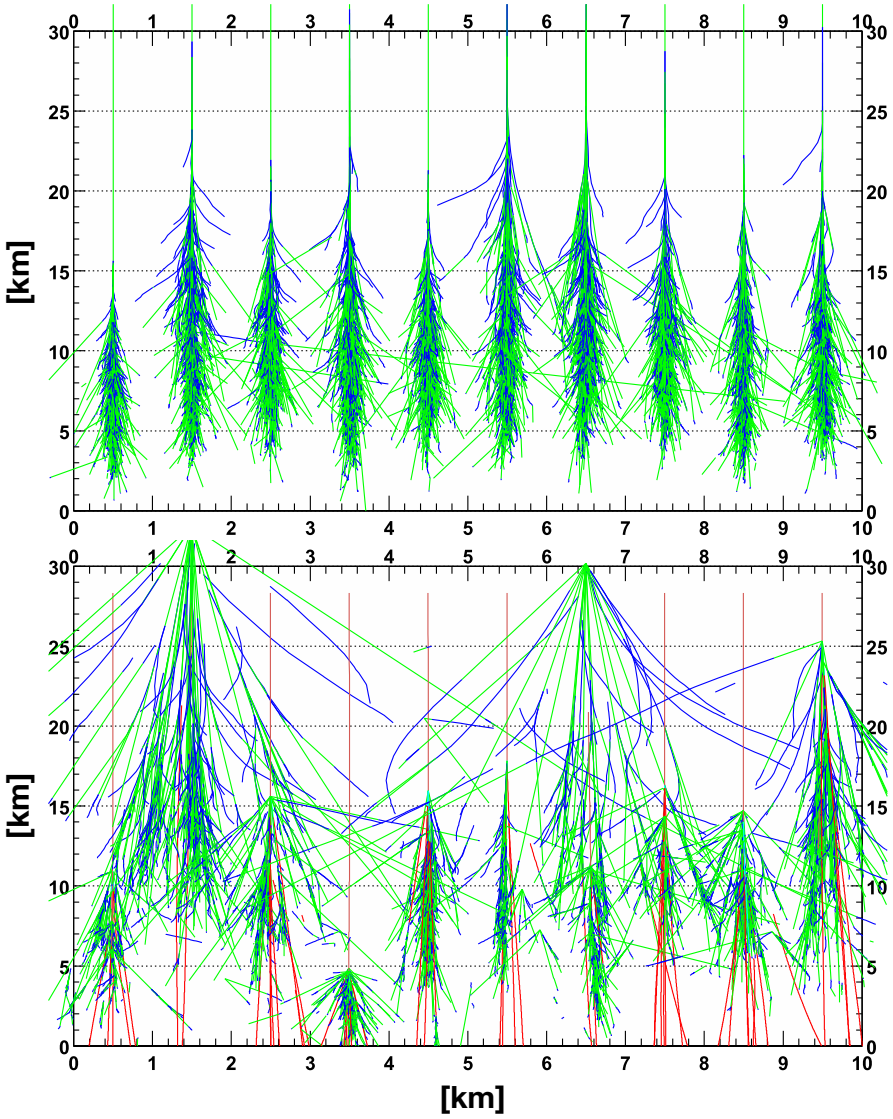


Figure 2.9: Variability of shower developments for primary γ-rays and protons

Illustration of the intrinsic variability of shower development, showing 10 individual vertical shower simulations induced from γ-rays in top panel and protons in the bottom panel. The energy of primary particles is 300 GeV. The showers are shown at the interval of 1 km next to each other. Figure extracted from [89].

critical energy E_c^{π}. Interacting charged pions generate again two-thirds charged and one-third neutral pions [104]. The neutral pions initiate sub-showers where Heitler's model can be adapted [78].

The discussion above can also be applied to heavier primary particles. One can consider a nucleus of mass A approximately as A independent nucleons, because the typical interaction energy is much larger than binding energy of 5 MeV. In this superposition model, a nucleus with mass A and energy E_0 is considered as A independent nucleons with energy $E_h = E_0/A$. [52]

Longitudinal and lateral distribution

The longitudinal development of a hadronic shower is more complex than that of a pure electromagnetic shower, because there are many more factors to affect it significantly.

For discussing the first interaction point, the behavior is different depending on the mass (species) of the primary. The hadronic interaction length is 70 g·cm^{-2} for protons and approximately 15 g·cm^{-2} for iron nuclei at PeV energies [102]. Therefore, the proton interaction length is longer compared to γ-ray and the first interaction point is lower in altitude. Correspondingly the first interaction point fluctuates more. On the other hand, a hadronic shower from heavier nuclei has a first interaction at higher altitude, with less fluctuation.

Subsequent development depends also on the primary particle. In proton case, the development is longer than heavier nucleon, because the behavior of a nucleus with mass number A is like the superposition of A showers with A times lower energy, as discussed above.

Comparing the maximum height of proton showers with that of an electromagnetic shower, they are comparable if the primary energies are the same. It would be explained by the net effect from the interaction length and the multiplicity. The interaction lengths of the participating squadrons are longer than the radiation length. The nucleon and pion interaction lengths in air are energy dependent, but good approximate numbers at energy below 10^5 GeV are 85 and 120 g·cm^{-2}, respectively. Therefore the hadronic component of the shower carries a large fraction of the primary particle energy deeper than electromagnetic cascade would. On the other hand, the secondary particles multiplicity in hadronic interactions is higher than the effective multiplicity of two secondaries in electromagnetic interactions and this faster energy dissipation compensates for the longer interaction lengths [104].

At atmospheric depths beyond shower maximum, there is little influence from the hadronic core, and the shower behaves like an electromagnetic cascade. The remaining particles continue to decay and the detectable particles on the ground will be mostly muons, which live long and many of them reach on the ground. The effect of muon cannot be neglected for ground-based detectors for EAS, and it is also the case for this study. It will be discussed in section 2.4.4.

Lateral distribution

In a hadronic shower the secondary particles created via hadronic interaction acquire large transverse momentum. Typically, the transverse momentum is almost energy-independent, $p_{\perp} \sim 350 - 400$ MeV, leading to a large angle of

low energy hadrons relative to the shower axis. In contrast, most of the electro-magnetic particles (i.e. electrons and positrons) are in the cascades initiated by high-energy π^0 nearly parallel to the hadronic core. Thus the lateral distribution of muon, which is produced mainly in the decay of low-energy pions, is wider than that of electro-magnetic particles [52].

2.3 Radiation from EAS

The charged particles of an EAS generate light due to two physical processes; namely, the Cherenkov effect and the fluorescence emission [88]. It means we can "see" the image of EAS through the emission from the shower. Because of the different emission mechanisms, the ways to make use of them are also different. In this study, the Cherenkov radiation is focused, and this section describes mainly this radiation and its effects, for the detection of γ-rays with the ground-based observations.

2.3.1 Cherenkov Radiation

Cherenkov radiation occurs when charged particles traverse a dielectric medium of refractive index n with a velocity $v = \beta c$ larger than the local phase velocity of light c/n. Figure 2.10 shows a sketch of the cases of if radiation occurs or not, depending on the velocity. If the velocity of the charged particle is below c/n, the polarization appearing near the track in the dielectric matter disappears immediately, thus the interference of dipole radiations from individual particles remains destructive. However, when $v > c/n$, the phase interference becomes constructive and coherence occurs.

The interference can be interpreted using Huygens law, in which the constructive interference appears as the superposition of the wave front that make a line on which the phase of the individual dipole radiation is equal. Therefore the propagation of the radiation has directionality at the certain angle with respect to the direction of the charged particle, which results in conical shape of the radiation. The angle of the direction with respect to the trajectory of the charged particle, θ_c, is

$$\begin{aligned} \cos\theta_c &= \frac{1}{n\beta} \\ \text{therefore,} \qquad & 1 > \cos\theta_c > \frac{1}{n} \end{aligned} \tag{2.3.1}$$

θ_c is called Cherenkov angle. The maximum of the angle corresponds to the upper limit of the velocity $v \to c$ and the minimum $\theta_c = 0$ corresponds to the minimum speed to emit Cherenkov radiation dependent on the refractive index n, called Cherenkov threshold. The typical Cherenkov angle and the Cherenkov threshold at various altitude in atmosphere is in table 2.1.

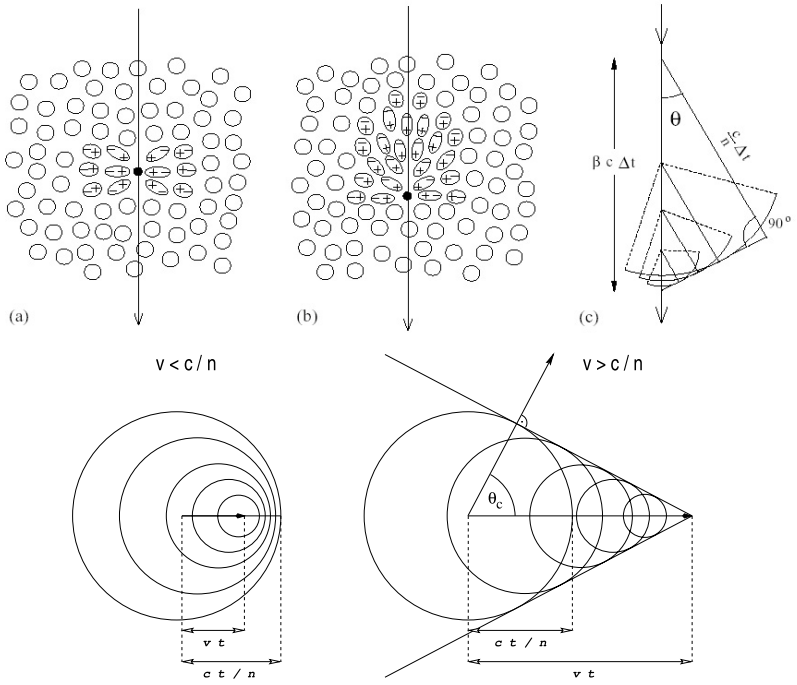


Figure 2.10: Cherenkov radiation

Top: The emission of Cherenkov light can be explained by investigating the polarization of the surrounding medium. (a) If a charged particle travels at a speed below the speed of light in the medium, a symmetric polarization occurs, and thus the dipole fields add up destructively. (b) If the particle moves at superluminal speed, the polarization becomes asymmetric and a resulting electromagnetic field forms. (c) Consequently, the electromagnetic field is enhanced by interference and the Cherenkov light wavefront is formed.
Bottom: Huygens' construction for determining the direction, in which the wavefront of the Cherenkov radiation propagates. In this way, Cherenkov light is interpreted as "shock wave" being produced by the particle moving at superluminal motion - in analogy to the well-known phenomenon of a supersonic shock front. A cone is formed, since only at this geometrical position the wavefronts interfere constructive.
Image taken from [65].

2.3.2 Light Pool

The directionality of Cherenkov radiation from EAS results in the limitation of the region to illuminate on the ground by the Cherenkov light. The illuminated region is called "light pool", as the examples shown in figure 2.11. The concentration of the Cherenkov photons within the light pool enhances the photon density in the illuminated region and is a big advantage to the measurement of EAS for better sensitivity.

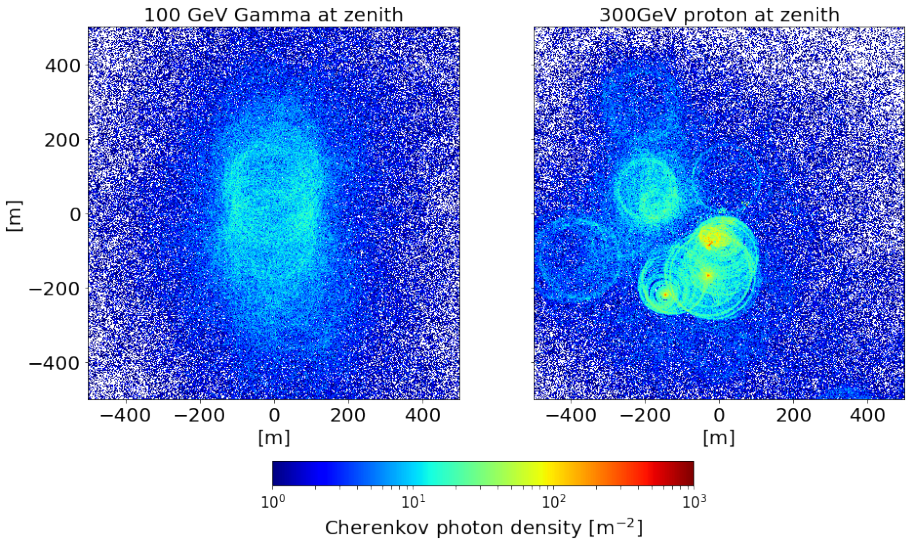


Figure 2.11: Examples of Cherenkov light pool

The two panels show the distribution of the Cherenkov photon density on the ground for MC simulated vertical showers initiated from a 100 GeV γ-ray (left) and a 300 GeV proton (right). Figure provided by Chai Yating.

A light pool is formed from an electromagnetic cascade, thus a γ-ray forms a single light pool. On the other hand a hadronic shower typically initiates multiple electromagnetic subshowers and forms multiple light pools. In addition, the hadronic component leaves muons which also form light pools in a different way (discussed in section 2.4.4). In an electromagnetic cascade, the Cherenkov photons are emitted from a number of electrons and positrons which have different energies thus different Cherenkov angles and have slightly different directions due to multiple scattering. Thus the photons emitted at an altitude make a disk of photons, expanding at the maximum Cherenkov angle as they travel downwards.

In the disk-like distribution of a light pool, there is a 'hump' at the edge (see figure 2.14). This comes from the varying refractive index dependent on altitude, as already shown in table 2.1. According to the Gladstone-Dale Law [105], the reflactive index $n(h)$ is a function of the local air density

$$n(h) = 1 + (n(0) - 1)\frac{\rho(h)}{\rho(0)} \tag{2.3.2}$$

where $n(0) = 1.000283$ [52] is the refractive index at sea level, and $\frac{\rho(h)}{\rho(0)}$ is

the ratio of air mass densities at the height h and at sea level, which can be calculated from a simple atmospheric model eq.(2.1.1). Correspondingly the Cherenkov angle changes and the photons from different altitude arrives on the ground after tracing different opening angles. The atmospheric parameters and the different trajectories of the outermost Cherenkov photons[10] are shown in figure 2.12. The figure shows that the region close to the edge is dense with the different trajectories, indicating accumulation of photons.

The figure also shows that a position on the ground can be illuminated by the photons from all the altitudes if it is close radial distance to the shower axis, while the farthest position is illuminated by the photons from only upper part of the shower.

[10] The upper limit of Cherenkov angle at the given height, assuming $\beta = 1$.

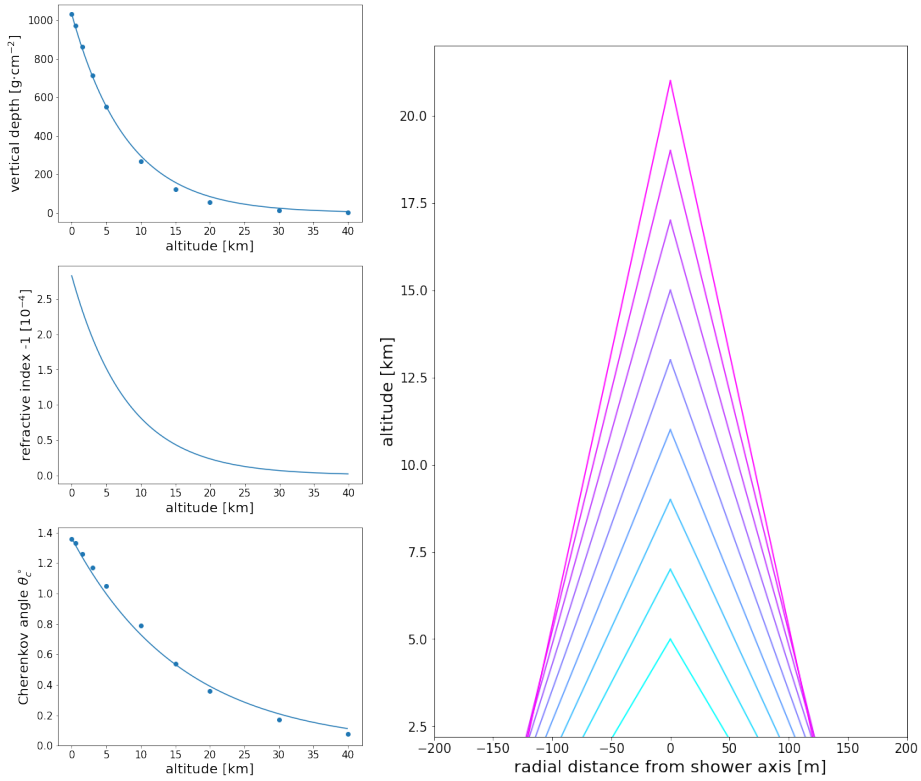


Figure 2.12: Atmospheric properties and characteristics of Cherenkov radiation

Left: The atmospheric parameters at different altitudes. From top to bottom, vertical depth, refractive index and Cherenkov angle. The points are the values shown in table 2.1, while the curves are calculated from eq.(2.1.1), eq.(2.3.2) and eq.(2.3.1), with scale height of 8 km.
Right: The lines show the trajectories of the Cherenkov radiation from a vertical EAS with the maximum Cherenkov angle, which is dependent on altitudes as $\cos\theta_{c,\max} = 1/[n(h) \cdot c]$. The different colors are just to discriminate the trajectories in the dense region. The emission altitudes are sampled every 2 km within the typical height to have significant Cherenkov emission from an EAS with primary energy 1 TeV. The values are calculated from the model shown in the left plots. The bottom of the vertical axis is set to be 2.2 km (MAGIC telescope).

2.3.3 Light yield and radiated/absorbed spectra

The energy loss by radiation per unit frequency interval of the radiated photons per unit length of the path of the radiating particle is known as Frank and Tamm formula [70]

$$\left(\frac{d^2E(\omega)}{dld\omega}\right)_{rad} = \frac{(ze)^2}{c^2}\omega\left(1 - \frac{1}{\beta^2 n(\omega)^2}\right) \tag{2.3.3}$$

where ω is frequency of the radiated photons and l is path length of the radiating particle. Integration gives the total energy loss. The range of ω is limited because the refractive index $n = n(\omega)$ and it is limited as eq.(2.3.1). The energy loss by Cherenkov radiation is a few % of MIP[11] and is negligible in the cascade.

The number of photons produced per unit path length of the particle dl and per unit frequency interval of the radiated photons $d\omega$ depends on their frequencies and can be derived via

$$\left(\frac{d^2N(\omega)}{dld\omega}\right)_{rad} = \left(\frac{d^2E(\omega)}{dld\omega}\right)_{rad} \Big/ (\hbar\omega) \tag{2.3.4}$$

It is normally discussed in wave length $\lambda = \frac{2\pi c}{\omega}$, and using fine structure constant $\alpha = \frac{e^2}{\hbar c}$.

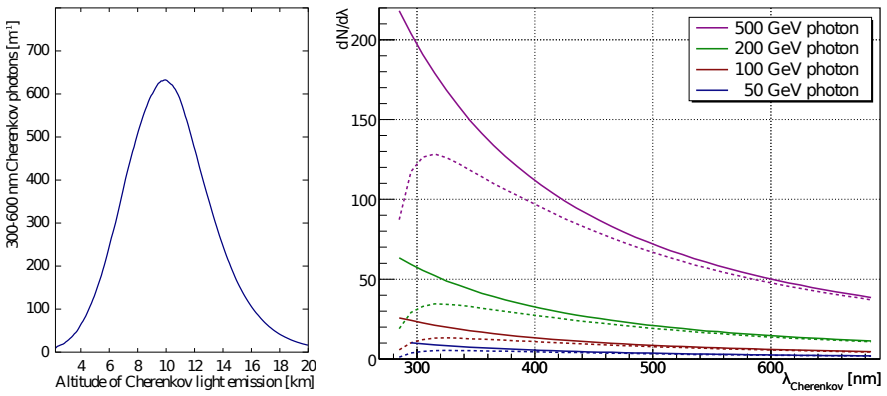
$$\left(\frac{d^2N(\lambda)}{dld\lambda}\right)_{rad} = \frac{2\pi z^2\alpha}{\lambda^2}\left(1 - \frac{1}{\beta^2 n(\lambda)^2}\right) \tag{2.3.5}$$

In Earth's atmosphere, n can be considered as a constant value for wavelengths longer than 250 nm which reaches the ground thus is important for ground-based observation. Therefore, the number of Cherenkov photons per unit wavelength interval is proportional to $1/\lambda^2$. Neglecting the dispersion, the integration over the wavelength λ_1 to λ_2 yields

$$\left(\frac{dN}{dl}\right)_{rad} = 2\pi z^2\alpha \sin^2\theta_c \left(\frac{\lambda_2 - \lambda_1}{\lambda_1\lambda_2}\right) \tag{2.3.6}$$

For an electron, the number of Cherenkov photons emitted between 300 nm and 600 nm is $\sim 10\ \mathrm{m}^{-1}$. As shown in the top left panel of figure 2.13, the total emission of Cherenkov photons from all the charged particles in EAS reflects the number of particles thus the production of photons is the largest at the shower maximum of ~ 10 km a.s.l.. Due to the absorption via different processes, the photons will be strongly absorbed, especially those with wavelength below 300 nm.

[11] MIP is a particle whose mean energy loss rate through matter is close to the minimum. The energy loss becomes close to MIP, when the velocity of the particle is relativistic, but not yet too high to have significant radiative losses. Such velocity range is, $\beta\gamma$ from about 1 to 100.



(a) (b)

Figure 2.13: The Cherenkov light yield and spectrum

(a) Average vertical light emission profile of a 100 GeV vertical electromagnetic air shower. Figure extracted from [27].
(b) The total spectrum of Cherenkov light produced by EAS, initiated by γ-ray primary particles with different energies. The solid lines are Cherenkov light spectra at 10 km height and the dotted lines are at 2.2 km above sea level (MAGIC telescope) after attenuation during the propagation of light. See figure 2.2 for the transmission as a function of the wavelength. Figure extracted from [113].

A clear atmosphere has good transmittivity down to 290 nm at which point ozone absorption sets in. Therefore, the most interesting waveband of the Cherenkov light from EAS to detect on the ground is from around 300 nm to several hundred nm. Figure 2.14 shows the expected density of Cherenkov photons on the ground (2200 m above sea level). For a location at the distance less than 100 m respect to shower axis, the expected number of photons m^{-2} is around 10 to 15, dependent on atmospheric profile.

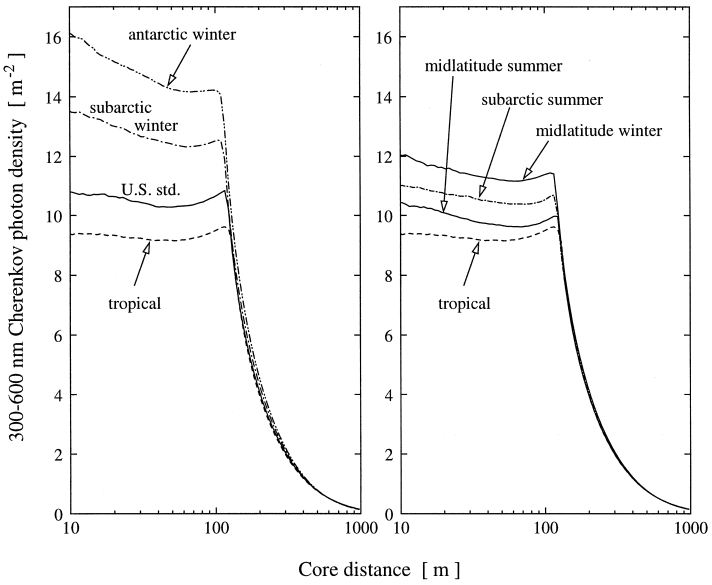


Figure 2.14: Radial distribution of Cherenkov photon density

Average lateral distributions of Cherenkov light photons in the wavelength range 300–600 nm for vertical 100 GeV γ-ray showers in CORSIKA 5.71 simulations with different atmospheric profiles (2000 showers simulated for each profile). The horizontal axis, core distance, is the radial distance from shower axis. The distributions are drawn in two panels to avoid the overlap of the lines and legends, while the distribution with tropical atmospheric profile is drawn in both panels. Absorption of Cherenkov light is taken into account. Observation altitude is 2200 m above sea level (the same as the MAGIC telescopes). Figure extracted from [27].

2.3.4 Arrival time spread

The arrival time on the ground of Cherenkov photons from EAS has typically several ns of spread, and the spread conveys the geometrical information of the EAS. There are two different effects that cause this spread. The schematic view of the effects are drawn in figure 2.15.

The first effect comes from the radial distance of observation point to the shower axis. Let us imagine that the shower develops at the same velocity as the speed of light and Cherenkov photons are still produced. The particles are always at the front of the shower, therefore the Cherenkov light front is formed always at the shower front, and there is no spread at the shower axis. On the other hand, the location far from the shower axis will have a timing spread, because the distance to travel for a Cherenkov photon from higher altitude is shorter than the distance of a path. For example, the trajectory of the Cherenkov light traveling straight from the top of the shower to an observer is shorter than the one, half of which is flown downwards as a secondary and another half is flown to the observer as a Cherenkov photon.

The second effect is induced from the refractive index. The charged particle in the shower travels faster than light. Therefore close to the shower axis, the

photons emitted at low altitude reach the detector before those emitted at high altitude.

By the two effects combined together, the arrival time spread has the specific trend dependent on the distance from the shower axis; the photons from higher altitude arrives later at near side to the shower axis, and vice versa.

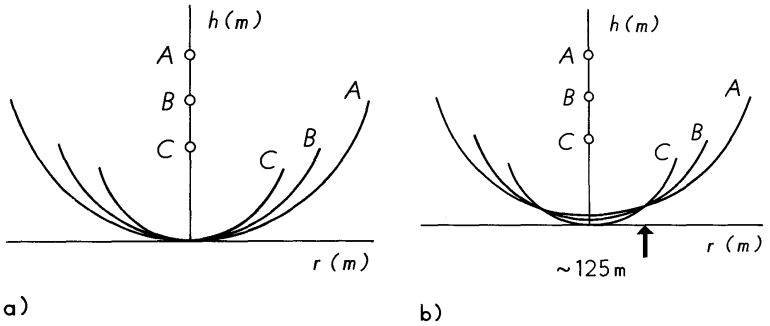


Figure 2.15: Arrival time difference of Cherenkov photons from different emission sites

A vertical shower is assumed, and the shower axis is drawn as the vertical line, along which three different emission heights (A, B and C) of Cherenkov light are indicated. The arcs indicate the shapes of Cherenkov light fronts from the three heights, (a) assuming no variation of atmospheric index of refraction and, (b) realistic case. The horizontal axis is the radial distance from shower axis. The arrow indicates the distance beyond which the index of refraction variation effects does not invert the order of the light fronts. Figure extracted from [102].

2.3.5 Effect of geomagnetic field

Without geomagnetic field the γ-ray showers look quite symmetric, and a light pool is formed in a circular shape on the plane normal to the shower axis. However the component of the geomagnetic field normal to the shower axis acts on the charged particles with the Lorentz force and bends the trajectories of electrons and positrons in opposite directions. It is called the east–west separation.

The separation leads to stretched distribution of light pool and the photon density in the light pool drops. It means less light content to detect, thus worse sensitivity.

The maximum influence is expected for EAS developing perpendicular to the direction of the geomagnetic field lines, i.e. for zenith distance (Zd) $\sim 39^\circ$ and Az $\sim 180^\circ$ at the location of the MAGIC telescopes [41].

2.3.6 High Zenith distance

When a primary particle penetrates the atmosphere at large zenith distance (Zd), the trajectory in the atmosphere becomes longer and the cascade reaches shower maximum at longer distance before the shower reaches on the ground.

Figure 2.16 shows the Cherenkov emission as a function of distance from observer on the ground, where the shower passes through. The peak height of Cherenkov emission changes dependent on Zd and the peak distance from observer becomes almost 100 km when the Zd is 80°, while it is only ∼ 10 km in case of low Zd.

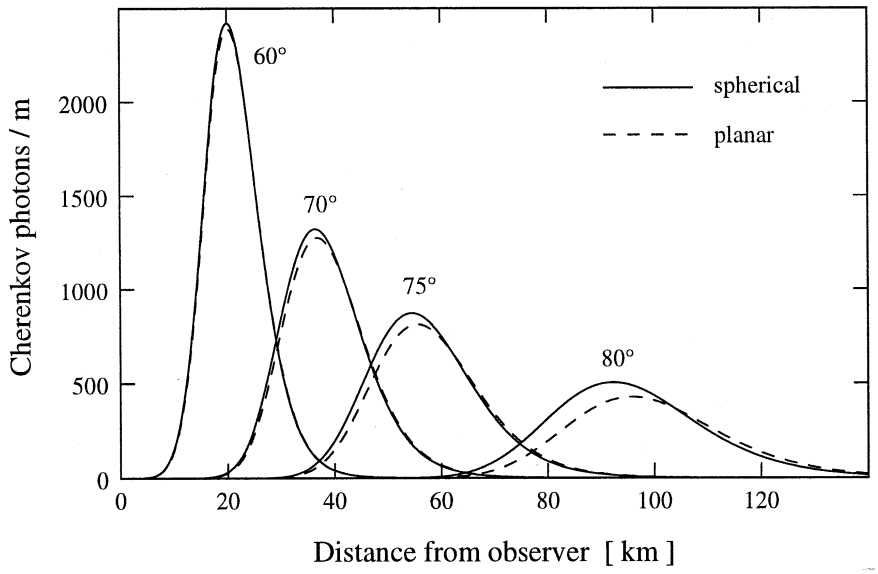


Figure 2.16: Longitudinal shower profile vs distance to observer for several Zenith distances.

The average longitudinal Cherenkov emission profile as a function of distance from the observer Cherenkov photons emitted in the wavelength range 300–600 nm per meter along the shower axis. An analytical approximation is used for different zenith angles (solid lines: spherical geometry, dashed: planar geometry). Figure extracted from [27].

2.3.7 Fluorescence light

Fluorescence is the emission of light by de-excitation of a substance that has excited by absorbing light or other electromagnetic radiation. In an EAS, fluorescence light is produced by the de-excitation of atmospheric molecules previously excited by charged particles of the shower.

Electrons and positrons in an electromagnetic shower passing through the atmosphere lose energy by inelastic collision with air molecules. A small fraction of the deposited energy is emitted as UV fluorescence radiation in the spectral range 290–430 nm. This air fluorescence in this spectral range basically comes from nitrogen [22]. The lifetime of the excited states of nitrogen is of the order of 10 ns. Therefore it is indistinguishable from Cherenkov light, however the contribution is negligible in the target energy range of VHE γ-ray because the fluorescence emission is isotropic. Fluorescence light contribution is significant when the shower energy exceeds $E > 10^{17}$ eV [52].

2.4 Imaging Atmospheric Cherenkov telescopes (IACT)

In the previous sections I explained the Cherenkov radiation from EAS, which is induced from a high energy particle including VHE γ-ray. This section introduces the basic concept of benefiting from the radiation, called the Imaging Atmospheric Cherenkov Telescope (IACT) technique. This technique is used to measure the image of the EAS via detection of its Cherenkov radiation, from which we estimate the incoming direction and the energy of incident γ-rays. This technique has excellent angular and energy resolution, with relatively large field of view. It also has good capability to determine the parent particle species thus to suppress the cosmic-ray-induced showers (gamma/hadron separation). There is now a general consensus that arrays of IACTs provide the most promising avenue for future VHE studies.

In section 2.4.1, the basic requirement for the ability of IACT is discussed, followed by the two important components, which are the reflector (section 2.4.2) and the camera (section 2.4.3). The basic properties of the images are described in section 2.4.4, and in the last, section 2.4.5 introduces an additional important technique to enhance the ability of the IACT, which is the stereoscopic observation.

2.4.1 Required ability

An IACT instrument is essentially an optical telescope to see the image of the EAS via detection of its Cherenkov radiation. The field of view needs to be wide enough so that the shower image at the most distant possible (the largest distance corresponds to the light pool radius, which is $\sim$ 120 m) is captured and it is typically $\geq 3.5^\circ$. It needs to overcome the very low brightness of the image[12] and high Night Sky Background (NSB) noise rate[13]. To deal with these problems, the telescope is equipped with a large reflector (typically in a short focal length f/0.7 to f/1.5 optical system) with a high-speed camera in the focal plane.

2.4.2 Reflector

The reflector is required to be large enough to collect enough photons. For typical IACT photon detection efficiencies (e.g. about 10% , because of the mirror specular reflectivity and the detection efficiency of camera), $\sim$ 100 m^2 of mirror area is needed to trigger a telescope with 100 photoelectrons[14] for a 0.1 TeV shower [51]. Current telescopes are based on either parabolic reflectors or the Davies–Cotton (DC) optical design. The parabolic design provides very good angular resolution on-axis, and moreover theoretically perfect timing conservation, but degrading resolution due to coma at larger field angles. On the other

[12] Cherenkov photon density is more or less constant within the light pool and can be approximated as 10 photons/m^2/100 GeV.

[13] $\sim 2 \cdot 10^{12}$ photons/m^2/s/sr

[14] In the condition that the incident flux of light is so low that detection of individual photons is required, the detection technique normally adopts photoelectric effect. The photoelectric effect can be attributed to the transfer of energy from the light to an electron. And the amount of electrons is measured via amplification to the current of charge. Thus the number of photons to be detected is measured in units of photoelectrons.

hand, the DC design reduces coma, the dominant source of off-axis aberration, providing a good Point Spread Function (PSF) over the entire Field of View (FoV), while the design introduces significant wavefront distortions (resulting in a time spread of several ns) and actually degrades the on-axis performance.

2.4.3 Camera

The image of γ-ray-induced showers seen from the ground looks like an ellipse with a typical root-mean-square width and length of 0.1° and 0.3° respectively, for a few-hundred GeV shower viewed at the zenith. To resolve the important structure of a shower image, the angular resolution of the telescope, angular diameter of the pixels and the pointing accuracy of the telescope mount should all ideally be $\leq 0.1^\circ$. However a pixelation in the camera much finer than this root-mean-square width does not seem to be very effective [51], because it degrades signal to noise ratio of a pixel.

Even after being collected by a very large reflector, the Cherenkov light of an EAS is still so faint that a pixel capturing the photons on the border of image needs to be sensitive to light down to single photons. Moreover, the signal needs to be discriminated from the background photons, which several ns of spread against the Poisson fluctuation in the night-sky. Therefore a pixel needs to adopt very sensitive detector with quick response. Currently the most utilized detector is Photo Multiplier Tube (PMT).

The electronics to record the signal is correspondingly required to minimize the integration time.

2.4.4 Imaging the EAS

The image of an EAS in the camera is just like a "picture", but with excellent time resolution, high sensitivity and enhanced brightness. Figure 2.17 shows a sketch of the relation between the EAS geometry and the image geometry in case of a γ-ray shower. The EAS from γ-ray can be seen as an ellipse in the camera, which reflects the longitudinal development of the shower along its major axis. The shower maximum, which is the brightest point of the image, is located around the center of the ellipse.

Figure 2.18 shows some examples of actual shower images of various EAS, triggered by MAGIC telescopes. The image of EAS initiated from a γ-ray is shown in figure 2.18a. The details about the images are described in the followings.

Effect of Zenith distance on the image

The image in the high zenith distance (Zd) observation is very different from the low Zd. The image becomes small (figure 2.18b) and the stereo images (introduced in section 2.4.5) become identical, because of the much longer distance from observer to the shower maximum with respect to the effective distance between the telescopes, as shown in section 2.3.6.

Effect of geomagnetic field on the image

Because of the geomagnetic field, the Lorentz force systematically deflects the electrons and positrons in EAS into opposite directions. This geomagnetic field

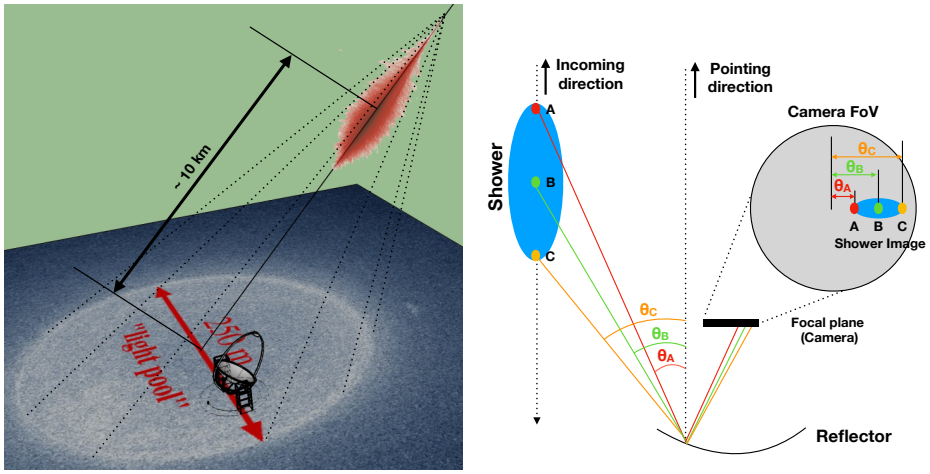
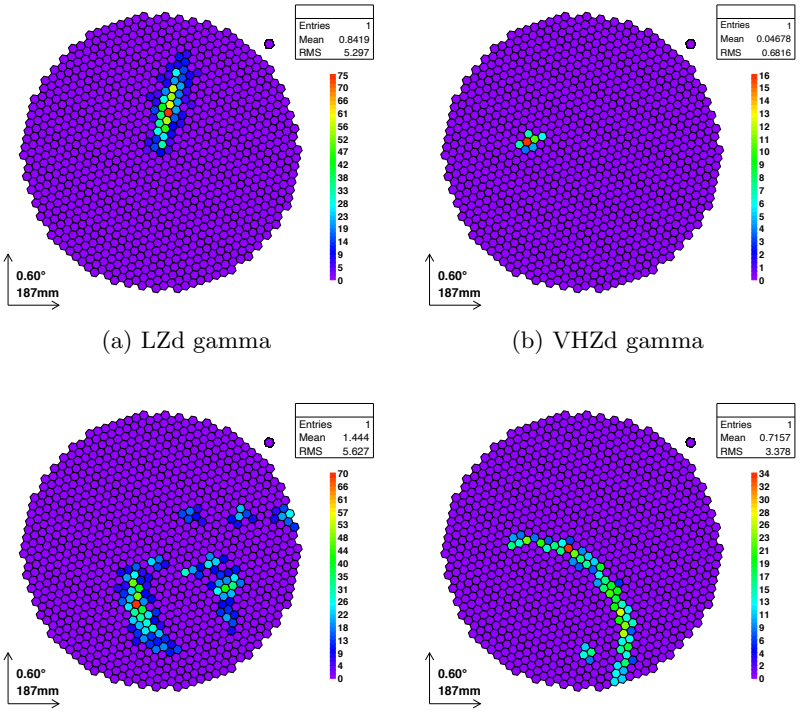


Figure 2.17: Sketch of the Imaging Atmospheric Cherenkov Telescope technique

Left: Sketch of a γ-ray shower, a light pool on the ground and the Cherenkov telescope. The scale of drawing is exaggerated. Typically a γ-ray shower develops at a distance of around 10 km if it comes from zenith, and the Cherenkov light arrives at the disk-like region with the diameter of 250 m.
Right: The example trajectories of Cherenkov light from three different radiation sites in the γ-ray shower through the reflector to the camera, and their correspondences to the shower image in the field of view of the camera. In this example the incoming direction of the γ-ray is assumed to be the same as the pointing direction of the telescope. The viewing angle between the direction of radiation site and the pointing direction or the γ-ray incoming direction is, from top of the shower (A) to bottom (C), small ($\theta_{\rm A}$) to large ($\theta_{\rm C}$). Reflecting this relation, the shower image is elliptical.



(a) LZd gamma (b) VHZd gamma

(c) LZd hadron (d) LZd hadron with muon ring

Figure 2.18: Various shower images in the PMT camera of a MAGIC telescope

Examples of images in the telescope camera capturing various EAS events. The images are cleaned; after calculating the number of arrival photons at the pixels, those which are dim and not clustered are removed.
(a),(b): Images of simulated γ-ray-origin showers when the telescope points towards low Zenith distance and large Zenith distance. The shower image can be seen as ellipses. The smaller image in the large Zenith distance is due to more distant location of the shower from the telescope.
(c),(d): The image shape can be used to estimate the particle species that originated the image of the EAS. The hadronic showers have irregular shapes. The large islands are normally from sub-shower components, the ring-like and arc-like shaped islands are from local muons, and the tiny islands may also be from local muons.

effect becomes non-negligible compared to the random displacement from multiple Coulomb scattering, and it degrades the measurement of very low energy γ-rays, lowering the trigger rate below a few 100 GeV by the reduction of the light content to detect.

The geomagnetic field also affects higher energy events in a relatively high Zd observation on the image shape. For EAS developing under unfavorable orientation with respect to the direction of the geomagnetic field, the corresponding Cherenkov light images in the camera of an IACT will be rotated [41].

Arrival time distribution in the image

As explained in section 2.3.4, the Cherenkov photons from EAS has the arrival time spread, and the spread has a trend reflecting the geometry of the shower, which is dependent on the distance to the shower axis (Impact). This trend is also seen in the image.

An image of EAS from a γ-ray is an ellipse, and the major axis of the ellipse indicates the shower axis. One side of the ellipse along the major axis is the top of the shower, while the other side is the bottom. Therefore the trend of the arrival times of photons appears along the major axis of the image.

As shown in figure 2.15, the arrival times of the Cherenkov photons from different emission heights are the same at around 125 m of Impact. When Impact is larger, the photons from the top part of the EAS arrives earlier than those from the bottom, and vice versa in smaller Impact. Consequently the difference of Impact appears in the arrival time distribution along the major axis in the elliptical image, namely the gradient of the arrival time along the shower image axis. This is called TimeGradient. This property of the image can be measured only by a parabolic reflector.

Hadronic images

The shapes of hadronic showers are typically far different from an ellipse, shown in figure 2.18c. The image consists of multiple clusters, which are called "islands". Typically, an island relates to an electromagnetic sub-shower component, in addition the image often contains islands from "local" muons.

Muon images

Among the secondary particles created in a hadronic cascade, a muon traveling at relativistic speed has a much longer life time and can reach the ground before decay. If such a muon passes through (or at least close to) the telescope, the Cherenkov light emitted at close distance is detected without losing intensity. Moreover, the Cherenkov light is emitted at the constant angle, which means the light comes to the reflector from the same viewing angle. The light is concentrated at the camera plane, thus the light from the muon creates a very bright image. It is called "local" muon.

The relation between some example trajectories of Cherenkov light and the images of local muons are shown in figure 2.19. When it passes through the reflector, the shape is a complete ring. The ring becomes an arc when it passes outside the reflector, then, the more distant, the more point-like. The point-like islands may appear similar to γ-ray origin. An example can be seen in figure 2.18d.

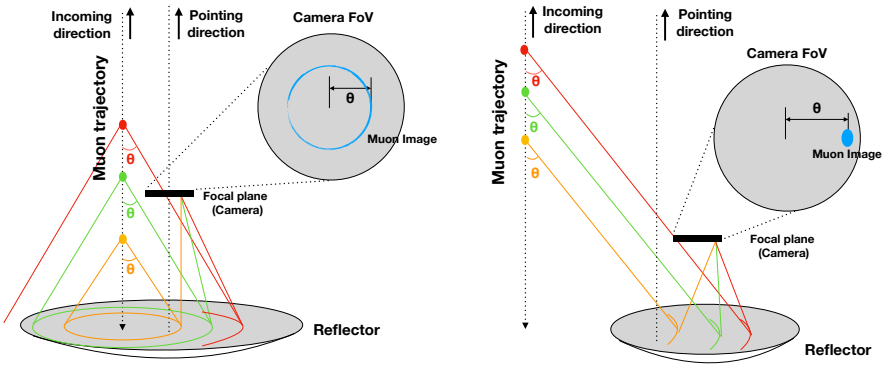


Figure 2.19: Sketch of imaging of muons in the telescope camera

Two examples of local muons, coming from the same direction as the pointing direction of the telescope. In the left panel the muon passes through the reflector, while in the right panel it passes nearby. The Cherenkov radiation from the muon (colored trajectories) is emitted with the Cherenkov angle $\theta \sim 1^\circ$, collected with the reflector, and detected at the camera.

Due to the large transverse momentum given to the secondary particles in a hadronic shower, the local muons sometimes arrive at a telescope without accompanying the Cherenkov light from other secondary particles or sub-showers of the same shower. Because of the risk of decreased effective time due to the pile-up of trigger dead time and of contamination of pseudo γ-like image in the analysis, it needs to be avoided. The stereoscopic observation (introduced in section 2.4.5) is useful also for rejecting muon-only events. If the trajectory of the muon is deviating from the electromagnetic subshower component, requiring coincidence between two a distance on the order of 100 m can work as an effective veto.

Due to the very clear behavior, the muon image can be used for measuring PSF and reflectivity. Related material can be seen in [65, 75].

2.4.5 Stereoscopic observation

In addition to selecting γ-ray events, the image of the air shower is also useful for estimating the properties of primary γ-ray. And the ability to determine the γ-ray properties is significantly enhanced by stereoscopic observation, which is the observation of the same EAS from different locations, shown in figure 2.20. In the compact elliptic shape of the γ-ray origin EAS, the major axis of the ellipse indicates the shower axis projected onto the image plane. Thus its origin (the incoming direction of the primary particle) lies along the major axis. This motivates the development of a stereoscopic system; if a γ-ray shower is observed by a stereoscopic system of two Cherenkov telescopes, its origin can be determined by superimposing the two images and intersecting their major axes.

Similarly it is possible to extract geometrical information of the EAS. Especially the distances of telescopes from the shower axis, Impact, is crucial information for the determination of the shower energy on the basis of the intensities

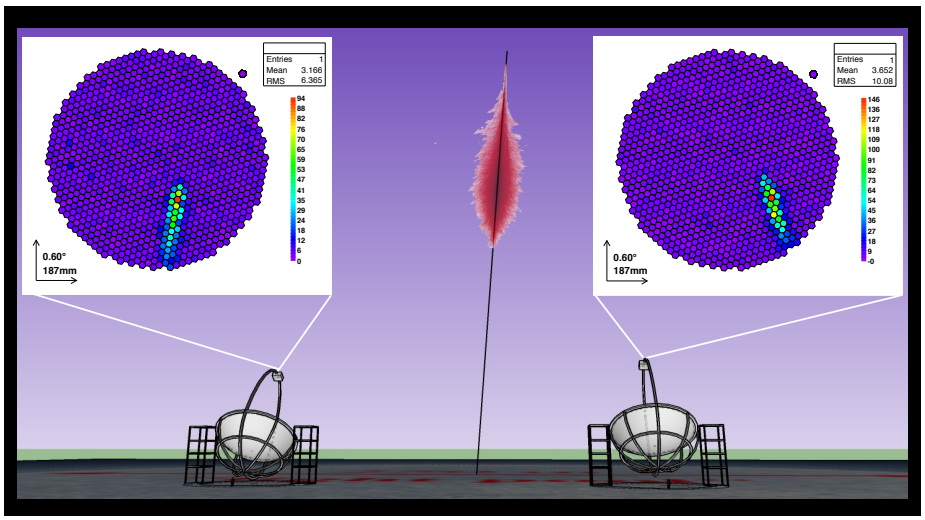


Figure 2.20: Sketch of stereoscopic observation of EAS from a γ-ray

In the sketch, two telescopes and an EAS from a γ-ray is drawn (red lines of particle trajectories) with the shower axis (black line). In addition, the two dimensional histograms are superposed, which are the camera fields of view. The histograms show the light contents captured by the pixels of the cameras in the two telescopes. The two telescopes point towards the same location in the sky, however the images look different, because the shower is seen from the different locations on the ground. However, the incoming direction of the primary particle is approximately the same in the field of view, because the incoming direction is far away.

detected in the telescopes [67]. The details are discussed in section 3.7.2.

The larger the separation distance of the telescopes, the better the stereoscopic images. However there is a limit due to the limited size of the light pool. Therefore the separation distance is typically around 100 m. Having multiple telescopes to see the same shower also contributes to background rejection.

The system needs deeper time depth, on the order of micro seconds, to temporally keep the signal before storing, to wait for the trigger signal for telescope coincidence to be propagated.

Chapter 3

The MAGIC telescopes and the analysis of MAGIC data

This chapter describes the MAGIC telescope system, together with the observation strategies and data analysis.

Section 3.1 introduces MAGIC, the observation strategy is explained in section 3.2, and the analysis method is described from section 3.3 on.

3.1 The MAGIC telescopes

MAGIC is a system of two 17 m-diameter Imaging Atmospheric Cherenkov Telescopes (IACTs) located at the Roque de los Muchachos Observatory (ORM) on the Canary Island of La Palma, Spain, at an altitude of 2200 m a.s.l (figure 3.1).

The telescopes were designed to reach the lowest possible energy threshold by optimizing various specifications; e.g. large reflector area, sensitive camera, high timing resolution and distance between telescopes. The telescopes are separated by a distance of 85 m from each other [17].

The achieved energy threshold is as low as ∼50 GeV at the trigger level for observations at zenith distance angles below 25° in dark conditions (without moon) [18]. Using the so-called sum-trigger, it is possible to reach an even lower energy threshold [21]. The benefit is not only the largest overlap in IACT with space telescopes like *Fermi* which covers lower energy range, but also ∼ 10000 times higher sensitivity than *Fermi*.

The second goal is to achieve a fast repositioning speed in order to catch rapid transient events such as Gamma-Ray Bursts.

Despite a few technical differences[1], the two MAGIC telescopes can be considered nearly identical for practical purposes.

[1] The two MAGIC telescopes started operation 5 years apart (MAGIC-I in 2004 and MAGIC-II in 2009, respectively), and the second telescope was an "improved clone" of the first one. There has been upgrades of the system in 2012 and 2013, which made them more similar to each other. [18]

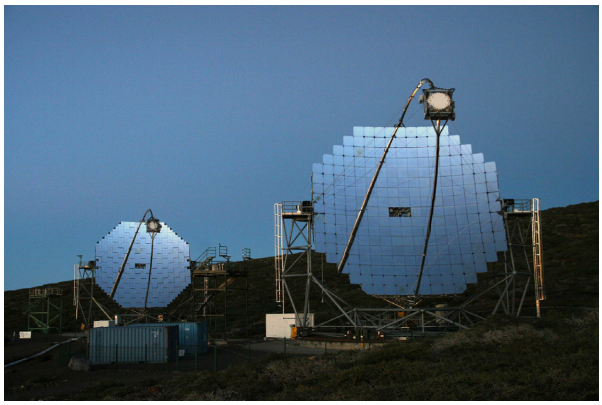


Figure 3.1: Photo of the MAGIC telescopes.
Image taken from [18].

3.1.1 Reflector, structure and drive system

Reflector

Each Telescope has a 17 meter diameter parabolic reflector system with focal ratio of f/1. The total area is about 240 m^2, which is composed of many segmented mirrors, attached on the individually adjustable mirror facets[2]. The PSF of the reflector[3] is $\sim$ 10 mm, which corresponds to $\sim$0.03°, which is less than one pixel. To approximate the parabolic reflector, the individual segmented mirrors have spherical surface with different curvature radii. Due to the parabolic shape, the time spread of synchronous light signals is negligible compared to the typical time spread of 1-2 ns in Cherenkov pulses. This means that the reflector does not introduce any significant broadening to the observed pulses.

Structure and drive

The structure and drive system are designed so that the telescopes can slew 7°/s, enabling fast repositioning to any direction in the sky within 20-30 seconds. To this end, the telescope structure is made out of reinforced carbon fiber tubes and achieved through a light weight ($<$ 70 tons) frame.

The camera is mounted on a vertical Aluminum arch that is supported against horizontal oscillations by steel cables. The telescopes are moved by two 11 kW electric motors on the azimuthal and one electric motor on the elevation axis [82].

[2] There are two sizes of the mirrors; 0.5 $\times$ 0.5 m^2 and 1.0 $\times$ 1.0 m^2. While the former is used exclusively for MAGIC-I, the latter is for MAGIC-II, as well as for upgraded mirrors in MAGIC-I. The adjustment of mirror facet is performed in order to keep the PSF of the reflector constantly to a minimum, by compensating the deformation of the support structure dependent on the pointing direction. The individual adjustment is available for single mirrors of MAGIC II and groups of 2 $\times$ 2 facets of MAGIC I.

[3] PSF of the reflector is the 39% containment radius of the reflected spot of a point-like source on the focal plane of the mirror.

3.1.2 Camera, trigger and readout system

Camera

Both MAGIC telescopes have cameras with 1039 pixels covering a 3.5° Field of View (FoV) [18]. Light detection in each pixel is done by a Photo Multiplier Tube (PMT). Using the photoelectric effect on photo cathode (the detection plane of PMT), the arrival photon stimulates a flow of bunch of electrons, which is enhanced using the dinodes with high voltages applied, to be visible as a pulse in probing voltage. Since the photo cathode of PMT has circular shape, the camera plane is covered by Winston Cones (WCs), whose outer shape is hexagonal to serve as light concentrators to circular inner shape. The shape of WC is optimized for collecting light only from the solid angle occupied by the mirror. The angular diameter mapped onto a single pixel is about 0.1°.

Trigger

Trigger system consists of multiple stages [18].

- Telescope level (standard trigger)

  The standard telescope level trigger of the MAGIC telescope system consists of two sub-levels, called "L0" trigger and "L1" trigger. L0 is a pixel-wise trigger to evaluate if the analogue signal from PMT is above a certain discriminator threshold. L1 is a digital coincidence trigger logic to allow only the moment when more than a required number of pixels next to each other fire L0 triggers.

  The threshold of L0 trigger is calibrated to take into account the difference of response by PMT, and the threshold level is adjusted online with respect to the Night Sky Background (NSB) level (trigger rate) of the observation direction, to avoid too high accidental trigger. L1 trigger is configurable for the number of neighboring pixels, which is called next-neighbor pixels (NN). It can range from 2 to 5, but usually 3 is chosen. L1 is performed only for the central pixels (figure 3.2).

- Coincidence of the telescopes

  The L1 trigger signals from two cameras generated are sent to the third trigger level, the stereo trigger (L3)[4]. The L1 signals are artificially stretched to 100 ns width and delayed according to the zenith and azimuth orientation of the MAGIC telescopes to take into account the differences in the arrival times of the Cherenkov light from air showers at the corresponding focal planes.

Readout

The current or voltage enhancement by the detection of photon occurs with time scale of ns. Thus very high sampling speed to record the varying voltage (waveform) is needed, and this is the core element of the readout system. MAGIC

[4] There was L2 trigger between L1 and L3, which performed rough analysis and apply topological constraints [45]. However it is not used now.

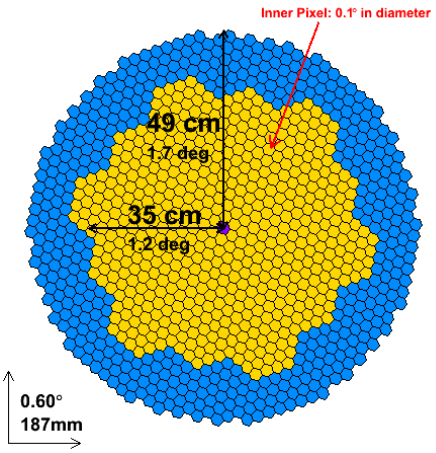


Figure 3.2: MAGIC camera and trigger region

The pixels in the trigger region are marked in yellow. Figure extracted from [75].

adopts the DRS4 waveform digitizer chip[5]. The DRS4 chip is a switched capacitor array, developed at the at the Paul Scherrer Institute that can store 1024 samples of the waveform and can be operated at sampling speeds of 0.7 to 5 GSamples/s. The sampling speed used in the MAGIC readout system is 1.64 GSamples/s to be able to look back to the time region which stereo trigger judges to take data[6]. The readout process itself happens at a much lower frequency and generates a deadtime of about 27 μs for each event. At a typical L3 trigger rate of 250 Hz, the deadtime fraction stays well below 1% [108].

3.2 Observation with MAGIC telescopes

The observation of a γ-ray source with MAGIC telescopes must be designed taking into account the background events. Although most of the background events are removed in the analysis of the observation data, they still dominate the remaining events in most of the cases. Therefore it is impossible to obtain a pure γ-ray event sample, instead it is necessary to evaluate if there are significant amount of excess events in the target region, compared to some control region(s), where no γ-ray source is expected. Correspondingly, the measurement of the γ-ray flux is possible only through the measurement of the excess events.

The target direction, where the γ-rays are supposed to come from, is called ON-region and that for control region is the OFF-region. Below I discuss how to obtain the ON data and the OFF data. The strategies to compare the ON and OFF data to extract the γ-ray signal will be discussed in section 3.9 and section 3.10.

[5] https://www.psi.ch/en/drs

[6] The time region of interest is 1/1.64 GHz $\cdot$ 50 = 30.5 ns long out of the buffer length of 1/1.64 GHz $\cdot$ 1024 $\sim$ 624 ns.

3.2.1 ON and OFF observation

The most simple way to determine the background is to use, in addition to ON observations (pointing at the region with the source), dedicated OFF observations (pointing at a region where we do not expect to see any γ-ray source). However this reduces the total amount of available observation time. In addition, as ON and OFF are not taken simultaneously in this strategy, it can be challenging to maintain the same observation conditions. The flux of background events changes mainly due to the change of pointing direction, as the source needs to be tracked. In addition, the NSB noise level, the transparency of the atmosphere, and the response of the camera (including the trigger threshold) also vary with time.

3.2.2 Wobble mode observation

This is the most common strategy to maximize the observation time as well as to reduce the systematic effect on the background estimation.

The source direction is slightly offset from pointing direction, namely the center of the field of view. The offset angle is called "wobble angle" (0.4° in MAGIC observations). The benefit of this strategy is the gain of OFF regions simultaneously at the same distance from the camera center, because the acceptance of the camera degrades radially towards the outside and is symmetric.

However, this can still introduce systematic differences caused by intrinsic inhomogeneities (e.g., slightly different Quantum Efficiencies of the PMTs in the camera or different noise level due to existence of some stars). To avoid this effect the pointing position is regularly switched (about every 20 minutes) such that the source position in the camera changes around the camera center.

The simplest switching is to choose a pair of wobbles, which have the pointing directions (camera center) offset with respect to the target direction in opposite sides (180°). The two OFF regions are towards different directions in the sky, however the background flux is assumed to be the same for almost the same directions.

The wobbling strategy can be elaborated more. Given the symmetric acceptance profile, the OFF region can be increased in the circle at the radius of wobble angle. Correspondingly the systematic effect should be reduced, by switching the pointing positions with respect to the choice of OFF regions. Figure 3.3 shows the example of 4 wobble positions. The offset directions of pointing positions are selected by every 90°.

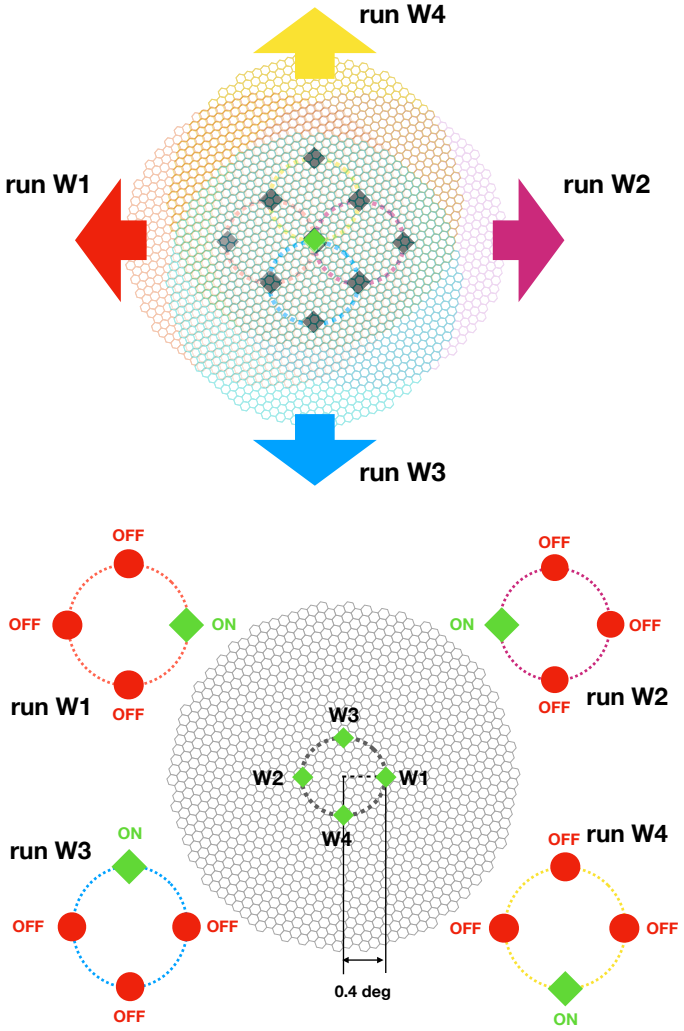


Figure 3.3: Schematic view of the wobble pointing mode.

An example of wobble pointing mode, with four wobble positions with 0.4° of wobble offset angle. As the top figure shows, there are 4 types of runs performed in turn, in which the pointing direction is offset from the target (γ-ray source) direction, shown as the green diamond, by 0.4°. The four regions in a camera plane, shown as diamond, are used for the comparison between the target direction (ON) and the direction without expectation of γ-ray flux (OFF). As the bottom figure shows, the ON region appears on different positions in the camera, shown as green diamond, over four consecutive runs (W1, W2, W3, W4). The other three regions are used as OFF regions for the background estimation.

3.3 Analysis chain with MARS

Almost all the analysis procedures of the observation data are carried out using the standard analysis package developed by the MAGIC collaboration named MAGIC Analysis and Reconstruction Software (MARS) [19, 116]. This software uses C++ routines combined with ROOT libraries [38].

In the following sections, I describe the MARS software in the phases of data analysis summarized as below.

section 3.4: Monte Carlo simulation We need a set of reference γ-ray events of the observations. However, a test γ-ray cannot be injected into the atmosphere to get a calibration of the whole system. Thus the reference γ-ray data are obtained via Monte Carlo simulations.

section 3.5: Photon signal extraction The observation data is comprised of many events in which the waveforms of all the pixels of cameras recorded at the readout level are stored. The first step of the analysis is to search in the waveform for the signal induced by the photons hitting the PMT. From the characteristics of the signal, the charge and timing of the signal are extracted. The charge will be interpreted as the unit of number of photoelectrons (phe).

section 3.6: Image cleaning The distributions of charges and their arrival times over the camera show the image of the EAS. But many pixels are dominated by NSB noise. These pixel information must be removed before the next analysis step.

section 3.7: Event parametrization The images are parametrized. And based on the image parameters, the geometrical parameters are also calculated.

section 3.8: Event reconstruction Using the parameters of the events, the property of the primary particle is estimated for each event. The event classification estimates the particle species, the direction reconstruction estimates the incoming direction, and the energy reconstruction estimates the primary energy before the EAS was initiated. The latter two estimations are done under the assumption that the primary particle is a γ-ray.

section 3.9: Extraction of γ-ray excess counts As discussed in section 3.2, the γ-rays can be extracted only by the number of excess events by comparing ON and OFF regions (observations). When the existence of γ-rays is not certain, the number of excess events must be compared with the background events and the significance must be evaluated.

section 3.10: Flux estimation The flux is an important measurement for a γ-ray source. Rich statistics allows detailed flux estimation, decomposing in the energy domain or time domain, which are called energy spectrum or light curve, respectively.

section 3.11: Spectral unfolding If the flux is investigated as a function of energy, the effect of confusion in the energy estimation needs to be taken into account. The distribution over energy is distorted by the confusion, therefore the correction process is needed. This is called spectral unfolding.

3.4 Monte Carlo Simulation

3.4.1 Purpose of the simulation

It is necessary to develop a detailed model of the telescope performance, as well as the whole process of the EAS from the primary γ-ray penetrating the atmosphere to the Cherenkov photons arriving on the ground. However we can not perform such calibration of the telescope using real γ-rays. Instead a chain of simulation programs is used. And a number of samples needs to be simulated in different configurations because the telescope response differs dependent on the EAS properties with respect to the telescope, like the primary energy of γ-ray, incoming direction, the arrival point on the ground, and so on. To reasonably cover the distribution of possible configurations, the samples are generated randomly. This strategy is called "Monte Carlo" (MC) simulation. There are two reasons why we need the simulated events of the γ-rays as a reference. The simulated samples are divided into two subsamples for respective purposes, which are called "train" samples and "test" samples.

"train" samples to estimate the nature of the primary particle
The simulated γ-ray events are used as reference to identify the real γ-rays from the observation data which is dominated by background events, as well as to estimate the direction and energy of the potential γ-ray candidate. The details are described in section 3.8. These estimations of the properties need the construction of the estimators, using the reference events, called training. The general explanation on the training and the evaluation of an estimator will be in section 4.1.

"test" samples to estimate the response of telescope and analysis
Once we obtain the γ-ray like events in the observation data, we need another set of simulated γ-ray events to infer the real flux from the observed distribution. The simulated γ-ray tells how much the probability of a γ-ray would be to survive the trigger system and analysis. The procedure to realize this will be described in section 3.10. Also, when the performance of estimator needs to be evaluated, this data set would be used. The explanation will be given in section 4.1.3.

3.4.2 Programs and strategies used in the MC simulations

The MC simulation for MAGIC [36] is performed outside the MARS framework. The simulation starts from a primary particle penetrating the atmosphere with a given energy, in a given direction and towards a given location with respect to the telescopes on the ground. It performs the generation of the secondary particles through interaction with the atmosphere, the generation of the Cherenkov radiation, and finally the detection with the telescopes. This is composed of three stages, associated to three different programs; the shower simulation part called CORSIKA, and the telescope part called Reflector and Camera.

1. "CORSIKA" (customized for MAGIC as MMCS): The simulation of EAS

   CORSIKA (COsmic Ray SImulations for KAscade) is a program for detailed simulation of extensive air showers initiated by high energy cosmic ray particles. It was developed for the KASCADE experiment and

is widely used for detailed simulation of air showers. Various kinds of primary particle like protons, light nuclei up to iron, and many other particles, including photons may be treated. For the MAGIC simulation, a customized version of CORSIKA is used, called MAGIC Montecarlo Software (MMCS).

The simulation starts from the cosmic ray primary particle. The important initial conditions to give are the primary particle species, the energy, the incoming direction (shooting direction), and the reaching point on the ground (Impact point), as well as the location information like the altitude of ground, the atmospheric density profile, and the geomagnetic field. In the simulation, a shower of the secondary particles are created from the primary particle through interaction with the atmosphere. All the secondary particles produced are tracked down to low energy until they produce no more Cherenkov light. The processes taken into account are as below.

- Hadronic interactions and decay of unstable particles
  The largest uncertainty of these simulations comes from the limited knowledge concerning the hadronic interactions at TeV energies. Since at the energies above TeV there is no experimental data on the cross sections yet, they have to be extrapolated from current measurements. Several models are in use to describe hadronic interactions and they are divided in low-energy and high-energy models. The transition energy between low and high energy models depends on the model combination and varies from 50 GeV to 1 TeV. In the standard MC package, the FLUKA model has been used, in combination with the high-energy QGSJet-II model.
- Particle tracking through the atmosphere
  Propagation of particles between two interactions accounts for ionization losses, deflection by the Earth's magnetic field and Coulomb multiple scattering.
- Electromagnetic interactions
  Unlike the hadronic interactions, electromagnetic interactions are few and well understood. The interaction of electrons, positrons and photons are simulated in CORSIKA using the EGS4 model. Bremsstrahlung, multiple scattering and annihilation processes are taken into account for electrons, while Compton scattering, electron-positron production and photoelectric reactions are considered for γ-rays.
- Cherenkov radiation
  In MMCS, the functionality related to Cherenkov photons is implemented. It calculates the production of the Cherenkov photons in air and the transport through the atmosphere. The program stores the information of individual photons reaching at the telescope; the arrival times, traveling direction and the height of production. The absorption of photons is not taken into account in this program, instead implemented in the program for later steps.

The program has been upgraded occasionally. MMCS currently integrates the version named Corsika 6.990 [64].

2. "Reflector": The ray tracing of the reflection at the mirror

   This program simulates the reflection of the Cherenkov photons by the mirrors onto the camera plane, as well as their atmospheric absorption and scattering before reaching the mirrors.

   - Atmospheric absorption
     For the atmospheric absorption and scattering of Cherenkov photons, the Rayleigh and Mie scattering and Ozone absorption are calculated. The emission height, arrival angle and wavelength of the photons are also taken into account. The atmospheric density profile is essential to the calculation, however it has seasonal tendency as well as daily or even short time scale variability. The standard production uses a fixed profile, called "MAGIC winter".
   - Ray tracing
     The photons survived the aforementioned process are passed to the ray-tracing of the reflection at the mirrors, and the program outputs the light distribution arriving at the camera plane. It will be passed to the next program called Camera.

3. "Camera": The detection of the photon at the Camera

   This program simulates the camera, readout electronics, and triggers of the telescopes. The arrival distribution of the photons on the camera plane is given to the pixels of the camera, and the responses of PMTs are generated as the waveforms of the electric current, if they pass the simulated trigger logic.

   All parameters of the telescope, such as gain fluctuation distributions, noise levels of the above listed components and signal shapes, are implemented in the simulation programs. These parameters are obtained from measurements.

After these processes, we obtain the events in almost the same format as observation data[7]. Each event consists of digitized waveforms from all the pixels in the camera of both telescopes. It can be processed with the data analysis in the same way as the observation data, mentioned in the next sections.

3.4.3 Standard data sets

A huge amount of simulation samples are needed because the telescope response to the γ-rays need to be understood in large variety; the γ-rays with different energy, incoming direction and arrival location on the ground, in different pointing configuration of the telescopes. Thus the MC simulation is standardized and the standard productions are commonly shared.

The simulated samples and real events should be ideally independent and identically distributed (i.i.d) in a given energy, given incoming direction, and in a giving telescope pointing.

[7] Observation events include additional information, e.g. the weather and the time. Contrary the simulated events include supplemental data like true information about the primary particle.

Parameter ranges

- Pointing direction (tracking)

  The standard MC production provides uniform distribution in Azimuth angle (Az) and Zenith distance (Zd), in given Zd ranges. However the production is separated by Zd ranges because the telescope response is different mainly due to the different energy threshold and collection area. The ranges are low Zd (LZd) defined as 5-35 deg and 35-50 deg, high Zd (HZd) as 50-62 deg and 62-70 deg, and very high Zd (VHZd) as 70-75 deg and 75-80 deg.

  Since the telescopes need to track a source, the analysis needs to take into account the telescope response dependent on Zd and Az. For Zd dependence, the telescope response is evaluated binned in Zd and the total response is calculated with weight to meet the observation time spent in the corresponding Zd (e.g. the calculation of collection area and energy migration. They are explained later). For Az dependence, the telescope response is not so significant in LZd, thus the usual analysis procedure calculates the averaged response over Az by taking all the events, however in HZd and VHZd the dependence on Az is significant thus the response is evaluated binned in Az.

- Incoming direction

  For a standard wobble-mode observation, the incoming direction with respect to the FoV is assumed to be offset by 0.4°, thus the events are simulated with the distribution of incoming direction in a ring of 0.4° radius (with a width of 0.1°) from the camera center. This is called "ring wobble" sample, and it is used for the analysis of point-like sources.

  On the other hand, for an analysis of γ-rays from unknown direction like morphology study, the telescope response must be evaluated also at any location of incoming direction with respect to the camera field of view. In this case, diffuse γ-rays are simulated covering a circle of 1.5° radius[8].

  Figure 3.4 shows the range of incoming directions of both types of MC.

- Energy

  The energy range of γ-rays needs to cover the range of possible detection. And it varies dependent on Zd range. For LZd, the range is from 10 GeV to 30 TeV. In the higher Zd range, the energy threshold becomes higher but the collection area becomes larger, thus the target energy range is higher. For HZd, the energy range simulated is from 50 GeV to 80 TeV for the Zd setting from 50 deg to 62 deg and 100 GeV to 80 TeV for 62 deg to 70 deg. For VHZd, it is from 400 GeV to 100 TeV for the Zd range from 70 deg to 75 deg and from 1 TeV to 100 TeV for the Zd range from 75 deg to 80 deg.

  Taking into account the balance between the low trigger rate in the lowest energy range, very heavy computational stress in the highest energy range, very wide range of the energy spanning in many decades and the real spectrum as power law, the generation of events follows power-law probability distribution with spectral index of -1.6.

[8] For full diffuse source there is also 2.5° radius diffuse source available.

(a) (b)

Figure 3.4: The incoming direction distributions of simulated γ-rays

The incoming distribution of the simulated γ-ray of the ring-wobble sample (a) and the diffuse sample (b). Image extracted from [81].

- Arrival point distribution on the ground

  The arrival point on the ground should be uniformly distributed in enough wide range so that effective area can be calculated properly (mentioned in section 3.10.1). The maximum distance to the telescopes is called max impact, where the trigger probability becomes negligible. The max impact is, for example, 350 m for LZd.

Analysis periods

There are versions of standard productions which are called "periods". A period is defined each time there is a major hardware intervention or a significant performance change due to weather conditions or minor hardware maintenance, so that MC simulations should accurately reproduce the real telescopes response.

3.5 Photon signal extraction

This step is performed with a program called *Sorcerer* (Simple, Outright Raw Calibration; Easy, Reliable Extraction Routines). The input data to this step is the raw data, in which, for each event, the waveforms of the electric voltage for all the pixels at the moment of trigger are stored. A waveform is a chain of measured voltage as a function of time, stored temporally in the capacitor arrays. The waveform stored in raw data is in the form of readout counts by Fast Analog to Digital Converters (FADCs), as a function of time slice (one time cell is 1/1.64 GHz). Figure 3.5 shows an example of such a waveform, which contains a signal of a light pulse in a pixel (PMT). The photoelectrons emerged from the photons hitting the photocathode of a PMT induce the flow of an amplified bunch of electrons, thus a pulse occurs in the time profile.

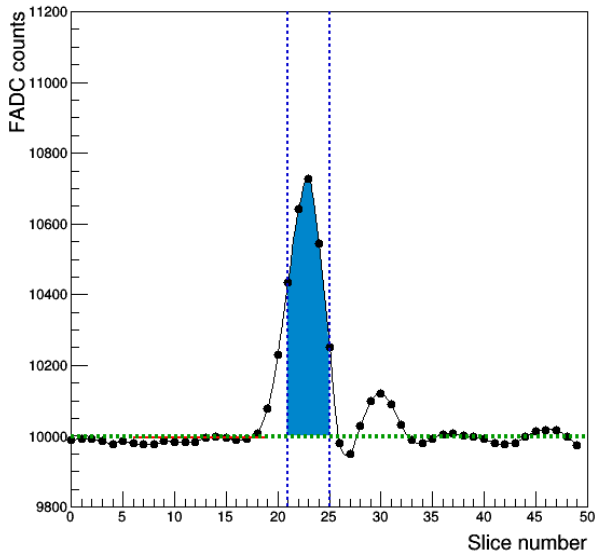


Figure 3.5: Signal waveform in a camera pixel

The signal waveform is sampled by the DRS4 chip (mentioned in section 3.1.2) at the frequency of 1.64 GHz, and 50 time slices (which is worth the time window of $\sim$ 30 ns) per event is saved. The vertical axis displays fast analog to digital converter (FADC) counts, which is linearly correlated to the current (and voltage) in the measurement range. The dotted green horizontal line is the baseline, and the blue shade surrounded by the two vertical dotted lines indicate the signal integration range. Figure extracted from [61].

Sorcerer searches for such a signal and extracts two pieces of information for the identified signal in each pixel in each event:

- Charge: The estimated number of photoelectrons detected
- Arrival time: Time at which this signal arrived

The charge can be measured by the current (voltage) integrated over time. However one needs to integrate the counts above the baseline in the region of the pulse, therefore the baseline and the integration region needs to be determined. For determination of the baseline, many events with random triggers, called "pedestal" events, are used[9]. They are taken before and during the observation. The integration region is a sliding window of 5 consecutive samples (i.e., the largest sum of 5 consecutive slices).

Then the conversion from the integrated FADC counts to number of phe is needed. Let the conversion factor be C, and assume n photoelectrons result on average in the charge $< q >$. Then,

$$n = C < q > \tag{3.5.1}$$

[9] The pedestal events should not contain any pulse inside. Therefore it is taken as the events with random trigger.

The conversion is not constant in time due to the performance of the amplifier during the signal transmission for data acquisition. Thus it needs to be calculated on-site using the so-called F-factor method [87]. For this method another bunch of events called "calibration" events is taken also before and during the observation. The calibration events are triggered at the moment when the calibration light illuminates the camera. The light is assumed to give a pulse of photoelectrons which follows a Poisson distribution with unknown mean N and standard deviation $\sqrt{N}$. The charge measured from the calibration events should be distributed correspondingly. Inspecting the distribution in the output of the PMT, namely the distribution of the integrated charge for the photoelectrons, the distribution is "diluted" from the intrinsic Poisson distribution. The modified relation between the mean and the standard deviation is kept independent from the conversion factor, then the mean and the standard deviation of the integrated charge $< Q >$ and σ_Q are in the following relation with the intrinsic Poisson distribution.

$$F\frac{\sqrt{N}}{N} = \frac{\sigma_Q}{< Q >} \tag{3.5.2}$$

Where F is called F-factor which is characteristic for individual PMTs, and can be measured at the lab. From the equations above, one can obtain

$$C = \frac{N}{< Q >} = F^2\frac{< Q >}{\sigma_Q^2} \tag{3.5.3}$$

Typically 1 phe corresponds to ∼60 readout counts integrated over 5 time slices.

As for arrival time, it is determined as the average of the time slices in the integration window, weighted with the signal heights.

3.6 Image cleaning

This step is performed with a program called *Star* [10]. The input data to this step is the charge (measured in the unit of phe) and arrival time for each pixel for each event. *Star* searches for the pixels which are supposed to be from the EAS, and removes all the rest of pixels.

If the pixels contain the photoelectrons related to EAS emission, they are supposed to be clustered spatially and temporally. Thus the algorithm picks up clusters of bright pixels and extends them to include dim pixels around the identified clusters, also taking care of the arrival times. Most pixel signals should contain only noise and such noise pixels are not clustered.

There are two stages in the algorithm. The first stage is to identify core pixels. It searches for groups of 4, 3 and 2 neighboring (4NN, 3NN, 2NN) pixels with a summed charge above a given level, Q_{core}, within a given time window, t_{core}. To prevent one pixel from dominating the signal of a group, the signal of each pixel is clipped with a given clipping level $Q_{core,clip}$. The second stage is to expand the core pixels. The selected pixels at this stage are called boundary pixels. It evaluates the neighboring pixels with charge above a given

[10] *Star* also performs different tasks, however I mention only image cleaning here. The other steps are image parametrisation for single telescope, and adding NSB noise to the pixels before the image cleaning. The former will be explained in section 3.7 and latter is in section 3.4.2.

level $Q_{boundary}$ and arrival time offset from that of adjacent core pixel within given time $t_{boundary}$.

The limit values above are dependent on the NSB level. The brighter the sky, the more the NSB noise, thus the higher the threshold needed. Depending on the condition, two values Q_1 and Q_2 are determined. In standard observation, $(Q_1, Q_2) = (6, 3.5)$[phe]. The values for core pixels are summarised in table 3.1, and the ones for boundary pixels are $Q_{boundary} = Q_2$, $t_{boundary} = 1.5$ ns.

Table 3.1: The limit values for core pixels search in the image cleaning algorithm

Topology	Q_{core}	$Q_{core,clip}$	t_{core}
4NN	$4 \times 1.2 \times Q_1$	$1.05 \times Q_1$	0.5
3NN	$3 \times 1.3 \times Q_1$	$1.05 \times Q_1$	0.7
2NN	$2 \times 1.8 \times Q_1$	$2.2 \times Q_1$	1.1

3.7 Event parametrisation

In this section I introduce the parametrisation of each event. The input data to this step are the charge and arrival time distributions over pixels after image cleaning. The distributions are parametrized in many different ways to deduce the properties of the incident particle of the event, namely the particle species, the incoming direction and the energy. Since the so-called "Hillas parameters" were proposed [66], the analysis with parametrisation have been the leading strategy. Although one can derive these properties without parametrisation via an elaborated machine learning algorithm, the estimation of the properties through the parametrisation works better for now. Moreover the parametrisation reduces the computational workload and gives transparency and robustness to the analysis.

There has been substantial amount of effort to improve the analysis performance, thus many parameters have been defined. Here I describe the image parameters that are used for the standard analysis, including the ones related to the γ-ray energy estimation.

The event parameters can be separated in two levels. One are the image-based parameters and the other are geometry-based parameters. The former are calculated in the program called *Star*, and the latter are in *Superstar* and *Melibea*. The image-based parameters need only the charge or arrival time distributions to calculate, while the geometry-based parameters are derived from incoming direction, together with the image-based parameters from the two telescopes.

3.7.1 Image-based event parameters

The main image-based event parameters are the Hillas parameters, which are CoG, Length, Width and Size. They are computed from the charge distribution of the image, through the first and second moments. In this computation, an ellipse can be defined using these moments, as well as its major and minor axes. They play the main role in the analysis, however there are some other complementary parameters.

CoG (MeanX, MeanY) The so-called "center of gravity" of the image. The CoG consists of a pair of values (X,Y) that determine the position in the camera of the weighted mean signal along the X and Y axis respectively. The X and Y values are the first moments of the charge distribution in the image and are called *MeanX* and *MeanY*, respectively.

Length, Width These are known as Hillas parameters [66], which characterise the image shape as an ellipse. Length and Width are calculated as second moment of the charge distribution along the major and minor axis.

Size The total charge, in the unit of number of photoelectrons (phe), summed over all pixels surviving the image cleaning.

Leakage (Leakage1, Leakage2) There are Leakage1 and Leakage2. Leakage1 is the fraction of the number of phe in the outer most pixels in the camera to the total charge of the image, i.e. Size. Leakage2 is the same for the pixels in the second row from the outer edge. In case the shower image is not fully contained in the Camera field of view, the image spans to the edge pixels on the camera field of view. Thus Leakage is correlated to the amount of spill-over of the shower image, and can help with correcting Size.

TimeGradient The arrival time development along the major axis [20]. The slope of the first order polynomial function, fitted to the distribution of the arrival time of pixel signals projected on the major axis.

NumIslands Number of islands in the image, namely the number of isolated groups of pixels that survived the image cleaning. Typically γ-ray showers produce only one island while hadron-initiated showers often contain two or more.

3.7.2 Geometry-based event parameters

The geometry-based event parameters are based on the three-dimensional reconstruction of the shower, which I call the geometrical reconstruction. The main part of the geometrical reconstruction is to determine the shower axis.

To determine the shower axis, two kinds of information are used; the incoming direction vectors and the shower core direction vectors of the two telescopes. They need to be determined from the shower images. The upper panel of figure 3.6 shows the shower axis in the relation with those information. The incoming direction of primary particle is approximately the vanishing point of the shower axis seen in the camera field of view. Therefore the incoming direction vector is approximately parallel to the shower axis. The shower maximum direction should be well represented by the brightest point, which is approximately CoG (image core).

There are currently two ways to estimate the incoming direction. One is by the classical method, which is the intersection point of the axes of the two superposed images. The other is by the Disp method, introduced in section 3.8.2. Note that an image axis does not represent the shower axis any longer in the Disp method (See section 3.8.2). Hereafter, I differentiate them as Classical-geometry-based variables and Disp-geometry-based variables. For the simulated

showers, the true geometrical configuration can be extracted. For those computed with the true geometrical configuration, I denote True-geometry-based variables.

Once the incoming directions and the shower core (the point on shower axis with shower maximum) directions are determined, the shower axis is determined in the following way. Since the shower axis can be seen as a line connecting the incoming direction and the image core in the camera field of view, the plane spanned by the incoming direction vector and the image core direction vector in 3D contains the shower axis. Thus the shower axis is derived as the intersection of the two planes created in such way by the images of the two telescopes.

It is also needed for some variables to obtain the shower maximum point. It should be well represented by the point where the two image core direction vectors meet, however it is almost impossible for the two vectors to exactly meet each other. Therefore the shower maximum point is defined on the shower axis, where the three points parallel to the ground, two on the image core direction vectors and one on the shower axis vector, make the smallest triangle.

I include Disp here, although it is not a geometry-based variable. It is neither an image-based variable. It is primarily to estimate the incoming direction, however it is also of great importance for the energy estimation. The Disp for direction reconstruction is mentioned in section 3.8.2 and that for the energy estimation is described in section 4.4.1.

Disp Disp is the angular distance in the field of view from the CoG to the incoming direction (see figure 3.9). It is not possible to measure in observation data, however the simulated γ-rays have the incoming direction of the γ-ray, which means that the simulated γ-ray has the true incoming directions, thus it has "TrueDisp". Disp is correlated to other event parameters, therefore it is possible to estimate. The Disp estimator is constructed using the simulated γ-rays, by supervising the relation between Disp and the correlated event parameters.

Impact The distance from the telescope to the shower axis. See also figure 3.6. Impact can also have true value in case of simulation data.

CosBSangle Cosine of the angle between the shower axis and the geomagnetic field. The component of the geomagnetic force normal to the shower axis act on the tracks of positively and negatively charged particles in the shower and bend them in opposite directions. Thus the shower particles are diluted and the Size will be reduced [41]. The shower axis is known in a simulated γ-ray thus the true value of CosBSangle is also available for the simulation data.

MaxHeight Height of the shower maximum point from the ground.

Cherenkov Radius / Cherenkov Density Cherenkov Radius is the radius, on the plane perpendicular to the pointing direction at the center of the two telescopes, of the Cherenkov light distribution produced by a 86 MeV electron[11] at the shower maximum. And Cherenkov Density is calculated

[11] An $e^{+/-}$ above this energy radiates high energy photons through bremsstrahlung, which then produce further $e^{+/-}$ by pair creation. A typical electron should have this energy at the maximum of shower development, because it is the point just before the number of particles start to decrease.

as the photon density on the plane in the Cherenkov Radius. They are measures to indicate the distance to the shower maximum.

Pointing Zd, Pointing Az The telescope pointing direction, in Zenith distance angle (Zd) and Azimuthal angle (Az). Characteristics of the shower images change significantly dependent on Zd and Az, especially at large Zd.

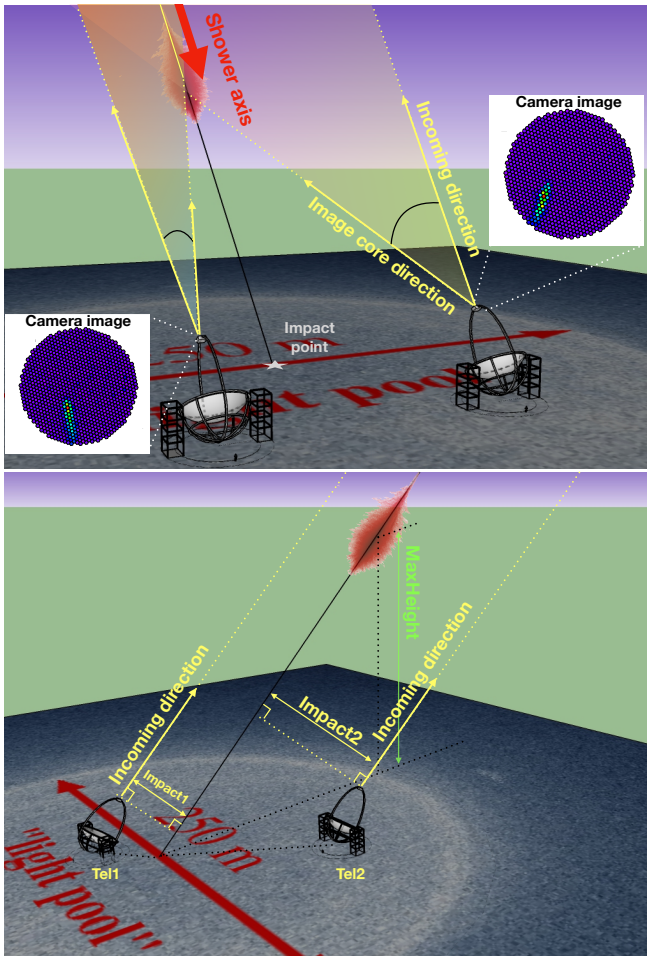


Figure 3.6: Sketch of shower axis, Impact and MaxHeight

Top: The shower axis in the relation with the incoming direction vectors and shower core direction vectors (represented by image core directions). The two vectors span a plane and the intersection of the two planes can be approximately regarded as shower axis. Note that the two image core direction vectors do note meet, hence additional considerations are needed to determine shower maximum point.
Bottom: The relations of Impact and MaxHeight with the shower axis.

3.8 Event reconstruction

This step is performed with two programs called *Coach* and *Melibea*. For studying a γ-ray source from an observation data, there are only three kinds of information needed about a shower event; the particle species (if it s a γ-ray or not), the incoming direction and the primary energy. Here I call these the event properties, and I describe the event reconstruction, namely the reconstruction of the event properties.

The three event properties are estimated from some of the parameters listed in the last section, by using the dedicated estimators. The estimators need to be constructed by supervisions in advance, for which we use simulated data, as discussed in section 3.4. The construction (training) of the estimators is performed by *Coach*. The estimation of the event properties using the estimators is done by *Melibea*.

Figure 3.7 shows an example of the event reconstruction. The reconstructed three event properties are shown; the particle species shown as Hadronness, the reconstructed direction shown as θ (the angular distance from the true incoming direction), and the energy. In the following subsections, I will describe these three event properties in detail.

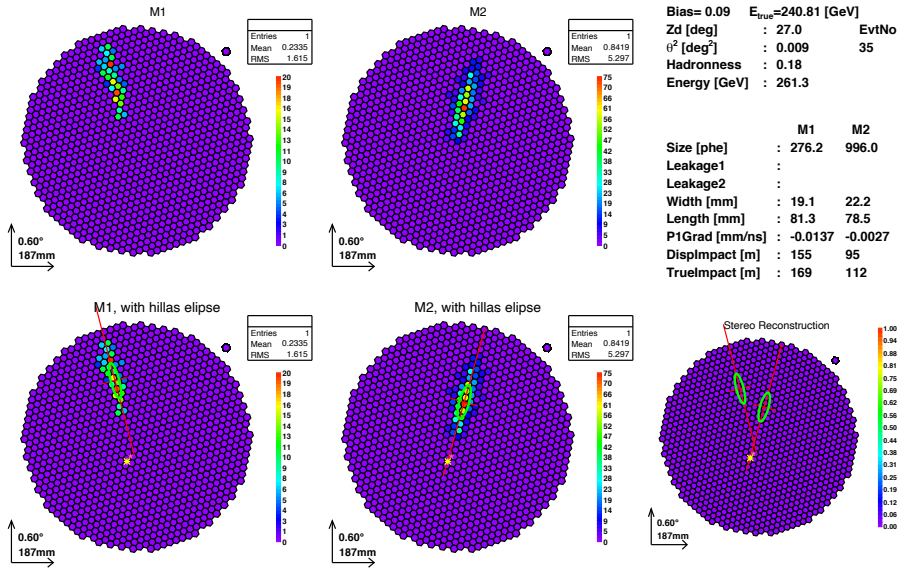


Figure 3.7: Example of event reconstruction

An example of event reconstruction for a simulated γ-ray event is shown.
Top right panel: The information of the event, which are the true values, the event parameters, and the reconstructed event properties.
Top left and top middle panels: The cleaned images of the γ-ray event captured by the two telescopes.
Bottom left and bottom middle panels: The cleaned images, superposed by the Hillas parameters (the green ellipse and the red lines as axes), reconstructed incoming direction (magenta star), and the true incoming direction (white star).
Bottom right panel: the Hillas parameters, reconstructed incoming direction, and the true incoming direction, without the shower image.

3.8.1 Particle species (Hadronness)

The observation data are contaminated with large amounts of background events, which mostly come from hadronic showers. In most of the cases, the background events dominate true γ-rays, and there must be some effective strategy to reject or at least reduce them in the analysis. To this end, the particle species is estimated. The background rejection by judging a γ-ray shower or a hadronic shower is called gamma-hadron (g/h) separation. Since the need is just binary decision, the variable "hadronness" is introduced. Hadronness is a measure to evaluate if it is γ-ray-like (Hadronness = 0 for a perfectly γ-ray-like event) or hadron-like (Hadronness = 1 for a perfectly hadron-like event).

Some of the image parameters show clear difference between γ-ray and the hadronic primaries, and they are used as the input variables to determine the hadronness. The computation of hadronness adopts Random Forest (RF). RF is introduced in section 4.3, and the details of the usage of RF for the hadronness estimator are given in section 4.3.3. The hadronness estimator is constructed using the supervision events, which are simulated γ-ray events as γ-ray shower examples and the observation data as hadronic shower examples.

The examples of the hadronness distributions are shown in figure 3.8. The observation data are dominated by hadronic events, thus its hadronness distribution peaks at 1. On the other hand, if a set of simulated γ-ray events is given, the distribution peaks at 0. Therefore, if a cut on Hadronness is introduced, a large fraction of hadronic shower events are discarded, while most of the γ-ray events are kept.

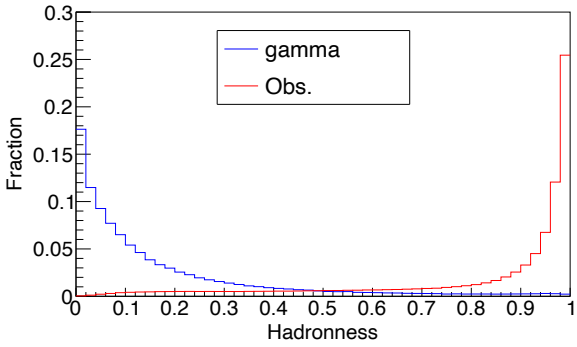


Figure 3.8: Hadronness distributions of simulated γ-rays and observation data

The distributions of Hadronness, assigned to individual events. Note that the data evaluated here are different populations of simulated γ-ray events and observation data, from those used as supervision samples for g/h RF.

3.8.2 Direction reconstruction and Disp method

In the standard MAGIC analysis, the event-wise direction reconstruction of the incoming γ-ray is performed with a DISP RF method (Disp method) [19].

The Disp method starts from the estimations of Disp for both telescopes, by a Random Forest (RF) regression (The explanation is given in section 4.3). Some

image parameters are correlated with Disp and the combination of them leads to the estimation. The Disp estimators are constructed for individual telescopes by giving the relations between the TrueDisp and the related parameters in MC simulated γ-ray events (the details of the usage of RF for Disp are explained in section 4.3.3. I also discuss the improvement of Disp in section 4.4.6).

Based on the estimated Disp, there appear two possible points as incoming direction, because the putative incoming direction is along the axis at the angular distance of Disp from the image core, in either side (left panel of figure 3.9). The problem is solved superposing the two images. When they are superposed, there are four pairs of possible combinations to indicate the incoming directions. The most probable incoming direction should be around the closest pair of the estimated directions.

If none of the four pairs give a similar arrival direction in both telescopes (namely the lowest distance is larger than 0.22°) the event is discarded. As the estimation of the Disp parameter is trained with simulated γ-rays, hadronic background events often obtain non-consistent results. Therefore this step provides an extra γ/h separation criterion.

The estimated incoming direction is computed as the average of the positions from both telescopes weighted with the number of pixels in each image. The angular distance from this point to the assumed source position is called θ.

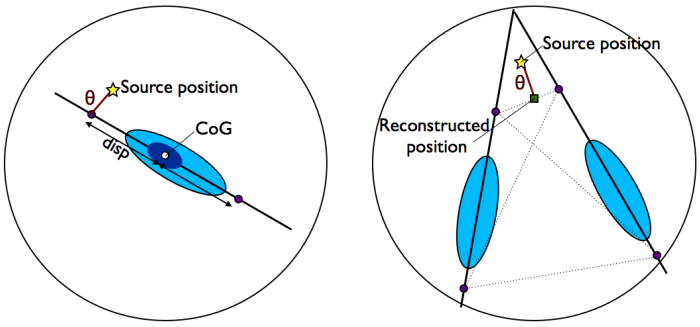


Figure 3.9: Sketch showing the principle of the Stereo Disp method.

The major axes of the images are plotted with solid lines. For each image, two possible positions shown as filled circles are obtained along the major axis (left image). However the true incoming direction, marked with a diamond, is not known thus it is impossible to choose one. When the images from the two telescopes are superposed (right image), the possible directions from the two images should be close to each other. The distances between the 4 pairs of the directions are evaluated, which are dotted lines connecting the filled circles, and the closest is chosen. The final reconstructed position (the filled square) is a weighted average of the two closest points. Figure taken from [81].

3.8.3 Energy reconstruction (LUTs-Erec)

The standard MAGIC analysis has adopted the Look Up Tables (LUTs) method for the energy reconstruction [17, 19]. This method has been used over the last decade, but now it is overtaken by my novel strategy, introduced in the next

chapter. Here I describe the LUTs method energy reconstruction (LUTs-Erec) in detail for a comprehensive comparison.

In the LUTs-Erec, the energy of an event is estimated by averaging individual energy estimators for both telescopes. This is because the Size, namely the light content observed, is a direct and independent information to indicate the primary energy of the γ-ray. The shower energy is approximately proportional to the number of electrons in the shower, as introduced in eq.(2.2.4), hence to the total light content emitted and to the Size of the image [116]. However Size needs corrections using additional parameters like Impact and Leakage to take into account the dependencies on the geometrical relation between the shower and the telescope. Therefore, if there are multiple telescopes recording the image of the shower, Size needs the corrections separately for individual telescopes at different locations.

Reflecting the thoughts above, the LUTs-Erec has three steps to reconstruct the energy. In the first step, several corrections to Size parameter are computed. Then, the energy of an event is estimated by two Look-Up-Tables for the individual telescopes. In the last step, the telescope-wise estimated energies are combined.

Size correction

Before the LUT is applied, the Size is corrected by empirical formulas to recover linearity of Size to energy.

- Zenith correction
  The larger the Zenith distance, the longer distance to the shower maximum point, thus the dimmer image results in smaller Size. To compensate this, Size is multiplied by
  $$\frac{0.97x^{-0.3}}{1-(1-x)^{2.25}}$$
  where $x = \cos Zd$.

- Leakage correction
  For the events with truncated images, Size needs to be compensated with the missing amount of light content outside the field of view. Size is multiplied by
  $$\frac{1}{1-4x^2}$$
  where $x =$ Leakage2.

- Magnetic correction
  In the presence of the geomagnetic field, the acceleration acts in opposite directions on electrons and positrons in the electromagnetic cascades. The separation leads to a stretched distribution of light pool and to a drop of light content to detect, as introduced in section 2.3.5. Size is multiplied by
  $$0.93 + 0.2\sqrt{1-x^2}$$
  where $x =$ CosBSangle.

- Max height correction
  The density of Cherenkov photons reaching the ground is affected by the distance to the shower. The density drop is characterised by the Cherenkov Density. Size is multiplied by
  $$\frac{1}{x}$$
  where x = CherenkovDensity.

Two dimensional Look-up table

After Size is corrected with all the factors mentioned above, the LUT is applied. The target value to search in the LUT is

$$\frac{TrueE}{CorrectedSize}$$

The LUT has two dimensions to search this value. One is Size and the other is Impact. Size is expressed as $\sqrt{log(Size)}$ for adjusting the grids, and Impact is expressed in the unit of Cherenkov radius so that it can take into account that the light pool radius changes according to the maximum height.

In the generation of LUTs, the target value in each grid is calculated from the train samples. They are filled to the histograms of the target value for corresponding grids, and the estimation value is determined by the peak position of the Gaussian fit to the histogram. For the combining process of the two estimations from individual telescopes, the uncertainty is also obtained as the width of the Gaussian fit. The application to each event is done by estimating the target value as the linear interpolation to the grids.

Note that the training of estimator (constructing the LUT) uses TrueImpact and TrueCosBSangle, which can be retrieved from the MC simulations. In the application to the data, the true information is not known and the estimated Impact (DispImpact) and CosBSangle (DispCosBSangle) are given as input to estimate the energy. I call this strategy as swap strategy. It is summarised in table 4.2.

Combining the estimation

The estimated energy of each event is the weighted average of both telescopes, where the weight is given by the uncertainty of the grid.

3.9 Extraction of γ-ray excess counts

Once the event properties are assigned, we can evaluate if the data sample contains a signal. As discussed in section 3.2, the statistical confirmation of the existence of genuine γ-ray source requires the comparison of the observed counts between the region around the direction of an expected source (ON-region) and the control regions (OFF-regions). The observed counts are derived from the so-called θ^2 histogram, and the computation of the significance from the comparison of the counts was formulated by Li and Ma [79].

3.9.1 θ^2 histogram

Figure 3.10 shows an example of real observation data, the first two runs (Run1 and Run2) of the GRB 190114C observation. The pointing direction is offset from the target direction by 0.4 deg, and is altered run by run. Top panels are the two dimensional histograms of reconstructed incoming directions of the events in the two runs. The ON-regions and OFF-regions are surrounded by the red circles and cyan circles respectively. The γ-ray signal is clearly visible in the ON-region of the Run1 data because of the extremely high γ-ray flux from this source, which decayed exponentially with time (see chapter 5 for further details).

The actual range of ON and OFF regions needs to be adjusted to minimize the background events. The γ-ray events spread around the target direction, thus the most efficient way is to define the circular region around the ON direction and adjust the radius. To this end, the θ^2 distribution is used.

The bottom panel of figure 3.10 shows the θ^2 distributions around ON and OFF direction. θ^2 is the angular distance to the target (ON or OFF) direction squared. For each event, θ^2 of ON and OFF are calculated and filled into the two histograms.

When the θ^2 histograms of ON-region and OFF-region are made from the two runs, the two histograms are expected to be consistent under null hypothesis, as the exposures in terms of the camera response should be the same in total. The figure shows a clear excess at low θ^2 values, which are the real γ-rays from this object. The determination of the radius of the ON and OFF region, namely the maximum θ^2, is described in figure 3.10.3.

3.9.2 Number of excess events

In the example above, the observed counts of the γ-rays is regarded as the number of excess counts, comparing the counts in the ON region with that in the OFF region.

$$N_{\mathrm{excess}} = N_{\mathrm{ON}} - N_{\mathrm{OFF}} \tag{3.9.1}$$

The excess counts, N_{excess} is essential information for a discussion on the target γ-ray source, and it will be used for the estimation of flux from the source. As discussed in section 3.2.2, there are different strategies of observations, and the exposure of OFF region(s) is not equal to that of ON region in those cases. Therefore some normalisation process is applied to the counts in the off region(s).

$$N_{\mathrm{excess}} = N_{\mathrm{ON}} - N_{\mathrm{normalisedOFF}} \tag{3.9.2}$$

In case of an ON and OFF observation, $N_{\mathrm{normalisedOFF}}$ takes into account the difference in observation time and observation directions. For an wobble mode observation, the multiple OFF regions can be defined in the camera field of view, thus the number of OFF regions is taken into account.

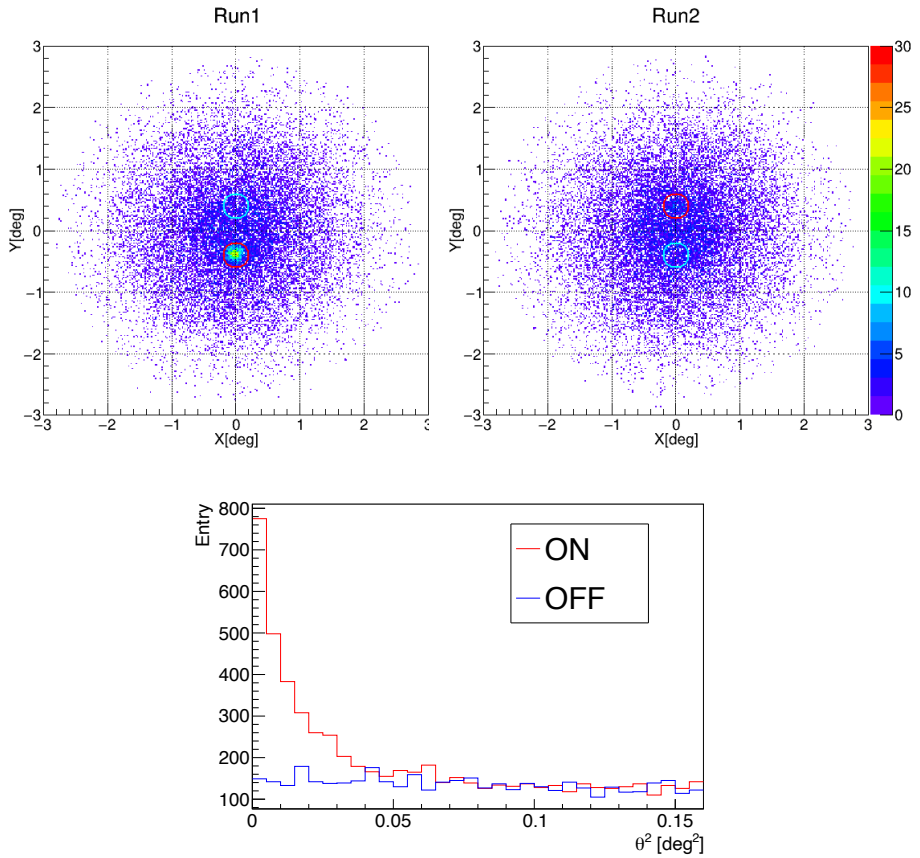


Figure 3.10: Example of the incoming direction distributions and θ^2 in ON and OFF regions

Top: The two dimensional histograms of the incoming directions, reconstructed for the two runs of observation data. The red circles are the ON regions, and the blue circles are the OFF regions.
Bottom: The θ^2 distributions around ON and OFF directions, accumulated from the two runs.

3.10 Flux estimation from excess counts

The main interest in a γ-ray observation is the γ-ray flux Φ, which is defined as the number of γ-ray particles passing through unit area per unit time.

On the other hand, the measured γ-ray events are the excess events that survived the trigger, the event reconstruction and the event selection. In the following, the derivation of flux is discussed, introducing the term "collection area".

3.10.1 Collection area (effective area)

The flux of γ-rays from a source reaching at the Earth is uniform, and the measured counts of γ-rays is the accumulation of the events at different distances from the telescopes. On the other hand, the probability for a γ-ray arriving at

the ground to remain after all the procedure through trigger and analysis is dependent on the distance to the Extended Air Shower (EAS) from the telescopes. Let us define "ground", a plane perpendicular to the incoming direction of the γ-rays, and define the coordinates on the plane to be (x, y). Given the probability to survive for the γ-rays until the excess counts as $p(x, y)$, the number of γ-ray events N obtained in an observation of a γ-ray source with flux Φ after observation time T [12] is,

$$N = T \int\int \Phi \cdot p(x, y)\, dx\, dy \tag{3.10.1}$$

$A_{\rm coll} = \int\int p(x, y)\, dx\, dy$ is called collection area. The collection area is strongly dependent on energy and the observation direction, thus it must be derived in a range of energy and direction with negligible variance.

By lack of an adequate calibrated high-energy γ-ray source in the sky, $A_{\rm coll}$ has to be derived using MC simulations. $A_{\rm coll}$ depends on the energy E of the incident γ-ray, on the telescope pointing zenith distance Zd (at high Zd also on the azimuth angle), on the incoming direction and also on the atmospheric conditions. For a simple case, like observing at low Zd in stable atmospheric conditions, it is sufficient to take into account only E and Zd. The simulation events must be generated uniformly in the given area $A_{\rm sim}$. Assuming that $N_{\rm sim}$ events are simulated, and $N_{\rm coll}$ events are collected after all the steps from trigger, through event reconstruction until event selection. Then, the collection area is

$$A_{\rm coll} = \lim_{A_{\rm sim}\to\infty} A_{\rm sim} \frac{N_{\rm coll}}{N_{\rm sim}} \tag{3.10.2}$$

3.10.2 Light curve and spectrum

There are two ways to express the flux in more decomposed way; light curve and spectrum. Light curve is the flux integrated over a given energy range and in bins of time. The other way is the spectrum, which is to show the flux as a function of the energy of the γ-rays. Since the collection area is strongly energy-dependent, the calculation is always done over sufficient number of energy bins. I begin with introducing the calculation of the spectrum. The introduction to the light curve follows as the flux integrated over energy.

The dependency of collection area $A_{\rm coll}$ on the observation direction is taken into account by averaging $A_{\rm coll}$ over the observed directions, weighted by the time spent. Hereafter $A_{\rm coll}$ represents the averaged collection area. In the case the energy dependence is taken into account, for $N(t, E_i)$ events collected in the time interval $t - \Delta t/2 < t < t + \Delta t/2$ and in the energy interval $E_{i,min} < E < E_{i,max}$, the flux is

$$\frac{d\phi(t, E)}{dE}\Big|_{E=E_i} \sim \frac{1}{\Delta t} \frac{N(t, E_i)}{A_{coll}(E_i)} \frac{1}{E_{i,max} - E_{i,min}} \tag{3.10.3}$$

This is called the differential flux. While the differential flux has the unit of number of particles per unit area, it is also useful to express the energy flux as

[12] The observation time T needs to be the effective time of observation, in which the dead time in the observation due to triggering and data read out is taken into account.

a function of energy, called Spectral Energy Distribution (SED). In this study I mainly show SEDs.

And the integral flux, for light curve, is,

$$\Phi(t) = \int_{E_l}^{E_h} \frac{d\phi(t,E)}{dE}\, dE \sim \frac{1}{\Delta t} \sum_i \frac{N(t,E_i)}{A_{coll}(E_i)} \tag{3.10.4}$$

3.10.3 Event selection

For robust estimation of the flux, an event selection needs to be performed. There are three kinds of selection criteria; Size cut and two efficiency cuts over θ^2 and hadronness.

Size cut

Size is primarily correlated with the energy of the γ-ray. And it is in the lowest Size range that is likely to have large systematic uncertainty because the trigger efficiency drops very quickly towards threshold. In addition, the background events overwhelm the signals and the statistical uncertainty is also large. Therefore the events with lowest Size needs to be avoided in the time and spectral analysis. In general, the cut value is decided as the peak of the Size distribution for the simulated γ-rays surviving the analysis. Most analysis use the cut Size > 50.

Efficiency cuts for θ^2 and hadronness

The parameters θ^2 and hadronness are used to reduce background events. The distributions of θ^2 and hadronness for γ-ray events and hadronic events change significantly over energy, therefore the selection criteria (cut conditions) are assigned in energy dependent. For robustness, the cut conditions are determined from the desired γ-ray efficiency, where a given percentage of simulated γ-ray events are collected after the cut condition. Typically, the efficiency values used are 75 % for θ^2 and 90 % for hadronness. The actual cut values are looser at lower energies.

3.10.4 Energy threshold

The energy threshold of IACTs is commonly defined as the peak of event distribution as a function of energy. Figure 3.11 shows examples in two Zd ranges, displayed by the differential number of MC γ-ray events that survived the event reconstruction with Size larger than 50. For low Zd range, i.e. Zd $< 30°$, the reconstruction threshold energy is $\sim$ 70 GeV. Note however that the peak is broad and extends to lower energies. Therefore it is also possible to evaluate the performance of the telescopes and obtain scientific results below this energy threshold.

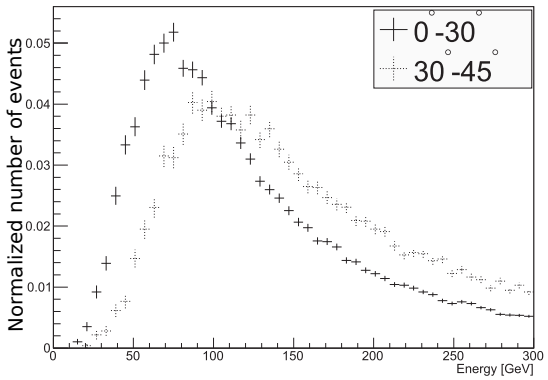


Figure 3.11: The event distribution used to determine the energy threshold

Normalized number of MC γ-ray events surviving the image cleaning with at least 50 phe for a source with a spectral index of -2.6. Solid line: zenith distance below 30°, dotted line: zenith distance between 30° and 45°. Figure based on the figure from [19].

3.11 Spectral unfolding and migration matrix

The calculation of spectrum described in section 3.10 needs a correction beforehand, which is called "unfolding", because the accuracy of energy reconstruction is limited and the mis-reconstruction (also called "energy migration") leads to deformations in the event distribution over energy. Figure 3.12 shows a sketch of such deformation.

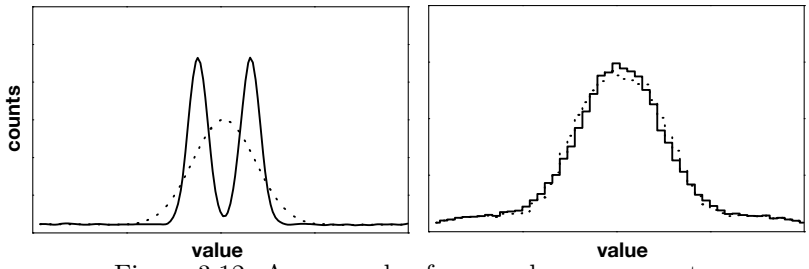


Figure 3.12: An example of smeared measurement

An example of smeared measurement. Two different distributions of some value with limited resolution become almost identical because of the limited resolution. Original distributions are solid lines and measured distributions are shown with dashed lines. Figure extracted from [32].

The deformation of the histogram binned in energy can be described as a matrix equation using the probability distribution of energy migration, called migration matrix (see also figure 4.35). Let us consider a histogram binned in estimated energy with n_y bins and a corresponding true energy histogram with n_s bins. Let the content of the estimated energy histogram be a vector $\mathbf{y}$, with an element for bin i as y_i. Similarly, the true energy histogram be $\mathbf{s}$,

with an element for bin j as s_j. While true energy distribution $\mathbf{s}$ is necessary for proper spectrum, the available distribution is only over estimated energy $\mathbf{y}$, which is deformed via convolution (or folding) with the migration matrix M. The migration matrix M_{ij} relates them as,

$$y_i = \sum_{j=1}^{n_s} M_{ij} s_j + \delta_i \tag{3.11.1}$$

where the matrix element M_{ij} gives the probability to find an event in bin i of the observed histogram which was produced in bin j of the true histogram, namely $\sum_i M_{ij} = 1$. And δ_i is noise term, which accounts for effects that are not described by the model of the measurement. The most obvious such effect is the statistical uncertainty, which comes not only from the fluctuation of signal events and their migrations, but also from background events, because $\mathbf{y}$ is obtained as excess events compared with control regions. In addition, δ_i may also account for the deviation of modeling the migration matrix.

To unfold the distribution, the migration matrix needs to be estimated. It is obtained from the MC simulation. Using the "test" samples, M_{ij} is estimated as [13]

$$M_{ij} = \frac{N_{ij}^{\mathrm{srv,rec}}}{N_j^{\mathrm{srv}}} \tag{3.11.2}$$

where N_j^{srv} is the number of events in j-th true energy bin which survived the event selection in the analysis, and $N_{ij}^{\mathrm{srv,rec}}$ is the number of events in j-th true energy bin which survived the event selection in the analysis and were reconstructed to be the energy corresponding to i-th estimated energy bin. Note that the distribution of the observation direction, especially in Zd, needs to be equalized between the number of generated γ-rays for the "test" samples and the duration of the observation time.

To solve eq.(3.11.1), a straightforward way is to approximate the equation neglecting the noise term and to solve it inverting M. However this computation tends to yield unphysical and extremely unstable results, and there needs to be more constraint for the solution. The strategy is to look for a solution close to a reasonable estimate of the spectrum (this is a requirement that concerns the space of solutions) and which gives a good fit to the data but without over-fitting (this is a requirement that concerns the space of data). The requirements in these two viewpoints cannot be simultaneously fulfilled and there needs to be a compromise. And there are two different approaches based on the difference in the weights on these two viewpoints; forward folding and unfolding.

In the followings, I explain the forward folding and unfolding in the MAGIC standard spectral analysis.

[13] Note that an energy bin has certain width thus the estimated matrix element is affected by the difference of the spectral slopes between the MC and the reality. To deals with this problem, the events are weighted for the spectral slope difference in a spectral analysis.

3.11.1 Forward folding

Forward folding assumes a particular spectral shape as a parametric function with a small number of free parameters, and it is fitted to the observed data. The fit process searches for the global maximum of the likelihood function, which is the joint probability density function of explaining the measured distribution. Not only the measured signal distribution $\mathbf{y}$, the background distribution can also be added to the function. Thus the likelihood function consists of the probabilities in each bin of the measured ON distribution $\mathbf{y}_{\rm on}$ and OFF distribution $\mathbf{y}_{\rm off}$, under the assumption of the signal distribution $\hat{\mathbf{y}}$ from the assumed spectral parameters θ and the background distribution $\hat{\mathbf{y}}_{\mathbf{b}}$ as nuisance parameters.

$$L(\mathbf{y}_{\rm on}, \mathbf{y}_{\rm off}|\theta, \hat{\mathbf{y}}_{\mathbf{b}}) = \prod_{i=1}^{n_y} P(y_{{\rm on},i}|\hat{y}_i + \hat{y}_{b,i}) P(y_{{\rm off},i}|\hat{y}_{b,i}/\alpha) \tag{3.11.3}$$

where a probability P is in Poisson distribution and α is the ratio of the observation time intervals in ON and OFF. To derive $\hat{\mathbf{y}}$ from θ, the spectral shape as differential flux is convoluted with the collection area as a function of energy and the migration is applied.

Forward folding requires reasonable spectral shape. Empirically and theoretically, most of the γ-ray spectra can be well described with power-laws, or power-laws with exponential cutoffs. If the source is at cosmological distance, the spectrum suffers from the deformation due to the energy dependent absorption of γ-rays. In that case, the forward folding can take into account the absorption in the fit process, starting the fit from the spectrum at the source (intrinsic spectrum).

The validity of the assumption is justified as long as the goodness-of-fit parameter (the probability to obtain the observed data if it is caused by the given spectral model) is reasonable.

3.11.2 Unfolding

Unfolding tries to perform a sort of inversion of the matrix M and derive $\hat{\mathbf{s}}$, namely the estimated distribution before deformation, without imposing a parametric functional spectral shape. However, the direct solution to the inversion problem tends to yield unphysical and extremely unstable results. Therefore there needs a regularization method, namely a reasonable assumption on the unfolded spectrum which gives a good fit to the data but without over-fitting. In this way, it attempts to filter out the noise in the unfolded distribution.

There are different regularization methods based on different assumptions. In MAGIC collaboration, there are five methods adopted [15]. The five methods are based on either the Tikhonov method [94], the Bertero method [28] or the Schmelling method [100]. These methods work in the way explained in either one of the two following approach bases.

Adding regularization term to χ^2 (Tikhonov and Schmelling)

The first approach is to search for $\hat{\mathbf{s}}$ with global minimum of χ^2. Assuming that the measured number of counts in the energy bins are independent from each

other with Gaussian distribution as an approximation to the Poisson distribution in large number, χ_0^2 is given, [14]

$$\begin{aligned}\chi_0^2 &= \sum_{i=1}^{n_y} \frac{(y_i - \hat{y}_i)^2}{\sigma_{y_i}^2} \\ &= (\mathbf{y} - M\hat{\mathbf{s}})^T \cdot K^{-1} \cdot (\mathbf{y} - M\hat{\mathbf{s}}) \end{aligned} \tag{3.11.4}$$

where K is the covariance matrix of $\mathbf{y}$.

The regularization is applied as the additional term into χ^2, which works as suppression on the overall shape of unfolded distribution.

$$\chi^2 = \frac{\omega}{2}\chi_0^2 + Reg(\hat{\mathbf{s}}) \tag{3.11.5}$$

where $Reg(\hat{\mathbf{s}})$ is the regularization term as a function of $\hat{\mathbf{s}}$, and ω is a weight which allows to steer the regularization strength. χ^2 minimum is searched changing ω. There are many formalization of the regularization term, and in MAGIC analysis adopts the Tikhonov method and the Schmelling method.

The Tikhonov method is the method of curvature regularization. The unfolded distribution is estimated by the smoothest or least curved distribution which is still statistically compatible with the measurements. The regularization term gives penalty on the frequent change of derivative, defined as

$$\begin{aligned} Reg(\hat{\mathbf{s}}) &= \sum_{i=1}^{n_y} \left(\left. \frac{d^2 s(E)}{dE^2} \right|_{E=E_i} \right)^2 \\ &\sim \sum_{i=2}^{n_y - 1} \left[2.0 \left(\frac{s_{i+1} - s_i}{s_{i+1} + s_i} - \frac{s_i - s_{i-1}}{s_i + s_{i-1}} \right) \right]^2 \end{aligned} \tag{3.11.6}$$

where in the middle term $s(E)$ is the $\mathbf{s}$ under the assumption that it is obtained as a continuous function of energy E. In the strong regularization, $\mathbf{s}$ approaches to smooth power-law distribution in general.

The Schmelling method is the method of reduced cross-entropy regularization.

$$Reg(\hat{\mathbf{s}}) = \sum_{i=1}^{n_y} p_i \ln \frac{p_i}{\varepsilon_i} \tag{3.11.7}$$

where p_i is the normalised distribution of $\hat{\mathbf{s}}$ and ε_i is a normalized prior distribution. The cross entropy quantifies by how much p deviates from ε. In the MAGIC standard analysis, the prior distribution is set to be the measured distribution, thus the unfolded distribution approaches the measured distribution itself in the strong regularization.

[14] In the χ^2 minimization of unfolding in the MAGIC standard analysis, the error of estimated value is added to the covariance matrix K, namely,

$$\begin{aligned}\chi_0^2 &= \sum_{i=1}^{n_y} \frac{(y_i - \hat{y}_i)^2}{\sigma_{y_i}^2 + \sigma_{\hat{y}_i}^2} \\ &= (\mathbf{y} - M\hat{\mathbf{s}})^T \cdot K_e^{-1} \cdot (\mathbf{y} - M\hat{\mathbf{s}}) \end{aligned}$$

Suppressing small eigenvalues of Grams's matrix

Another approach of regularization is to approximate the inversion of the migration matrix M, with suppression in the small eigenvalues. The matrix relation

$$\mathbf{y} = M\mathbf{s} \tag{3.11.8}$$

is solvable by introducing so-called Gram's matrix,

$$G = M \cdot M^T \tag{3.11.9}$$

G is symmetric thus is diagonalizable by orthogonal matrices, which can be denoted as

$$G = (\lambda_1 \mathbf{g_1}, \lambda_2 \mathbf{g_2}, \dots \lambda_{n_r} \mathbf{g_{n_r}}) \cdot (\mathbf{g_1}, \mathbf{g_2}, \dots \mathbf{g_{n_r}})^T = \sum_{l=1}^{n_r} \lambda_l (\mathbf{g}_l \cdot \mathbf{g}_l^T) \tag{3.11.10}$$

where $\mathbf{g}_l$ is the orthogonal vector with the correspondent eigenvalue λ_l. The solution of eq.(3.11.8) reads

$$\mathbf{s} = M^T (M \cdot M^T)^{-1} \mathbf{y} = M^T (G)^{-1} \mathbf{y} = M^T \sum_{l=1}^{n_r} \frac{1}{\lambda_l} (\mathbf{g}_l \cdot \mathbf{g}_l^T) \cdot \mathbf{y} \tag{3.11.11}$$

Thus small eigenvalues give strong oscillation in the solution. The Bertero's method tries to suppress the contribution from them by using the Richardson's iterative method. The equation is constructed for approximate solution of $\mathbf{c} = G^{-1}\mathbf{y}$ to iterate over $\mathbf{c}$.

$$\mathbf{c}_{i+1} = \mathbf{c}_i - \tau(G \cdot \mathbf{c}_i - \mathbf{y}) \tag{3.11.12}$$

where τ is the relaxation parameter. If $0 < \tau < 2/\lambda_{\max}$, the solution converges to G^{-1} for infinite i. In MAGIC analysis, $\tau = 1/\lambda_{\max}$. The solution depends on $\mathbf{c}_0$, the initial distribution of $\mathbf{c}$, and the number of iterations i. In the case of MAGIC analysis $\mathbf{c}_0$ is configurable to be either 0 or $\mathbf{y}$. In the case $\mathbf{c}_0 = \mathbf{y}$, the approximated solution of G^{-1} reads

$$\tilde{G}^{-1} = \sum_{l=1}^{n_r} \left(\frac{1}{\lambda_l} [1 - (1 - \tau\lambda_l)^i + (1 - \tau\lambda_l)^i \lambda_l] \right) \mathbf{g}_l \cdot \mathbf{g}_l^T \tag{3.11.13}$$

Search for the global solution

For the above mentioned strategies, the solution changes dependent on the weight ω for the former strategies or the iteration number i for the latter strategies. However there is no unique way to choose the best ω or i. Therefore in MAGIC analysis, the "best" value is searched over a wide range of ω or i, from 10^{-5} to 10^{10} and there are different criteria of choosing the "best" value. The options of the criteria are discussed in section H.1.

3.11.3 Pros and cons

Forward folding is very beneficial in the sense that a spectrum with theoretical model can be directly fitted to the observed counts in order to obtain best-fit

spectral parameters and a goodness of the fit. However, the fit never works if a reasonable spectral model is missing. In such a case, the spectrum needs to be reconstructed with minimum assumption beforehand so that the shape can be characterized. A spectrum can have various kinds of structure like a cutoff, a broken power-law or even a complicated irregularity.

To this end, unfolding is an indispensable alternative approach to the spectral analysis. A spectral shape can be reconstructed without any strict assumption on the shape, in as small scale as the energy resolution. However it does not take into account the fluctuation of background events. Moreover, it considers the energy migration only for the energy bins with enough statistics and the number of excess events are treated with Gaussian uncertainty. These treatments affect the low statistics region, namely at the edge of the range. In addition, the entire range can be affected in case of a steep spectrum, where a systematic uncertainty of spillover from lowest energies might change the slope.

Therefore both forward folding and unfolding should be cross-checked for consistency.

3.12 The performance of MAGIC telescopes

3.12.1 Crab Nebula as reference source

As discussed in the previous sections, the flux estimation of a γ-ray source is based on the telescope response measured using simulated γ-rays. Additionally, the overall performance of MAGIC observation and analysis is also evaluated using observations from the Crab Nebula. The Crab Nebula is a nearby ($\sim$ 1.9 kpc away, [110]) pulsar wind nebula, and the first source detected at VHE γ-rays [114]. At VHE γ-rays, it is the brightest persistent and not time-variable source. However, there has been a few γ-ray flares observed at GeV energies by AGILE & *Fermi*-LAT [4, 107] but no counterpart has been detected at VHE. Therefore, the Crab Nebula is commonly referred to as the "standard candle" of VHE γ-ray astronomy, and it is frequently used to evaluate the performance of VHE instruments.

The comparison of the spectra of the Crab Nebula drawn by different experiments and different versions of the MAGIC telescopes is shown in figure 3.13.

3.12.2 Sensitivity

The sensitivity of the instrument is calculated as the minimum flux level that can be detected at 5 σ of significance s, in 50 hours of observation. The significance here is calculated as $s = N_{\mathrm{S}}/\sqrt{N_{\mathrm{B}}}$, where N_{S} is the number of excess events and N_{B} is the number of background events.

The sensitivity can be calculated from the Crab Nebula observations. In general, N_{S} should scale proportionally with flux level in a given observation time, and N_{S} and N_{B} should scale proportionally with observation time t_{obs} in general, thus $s \propto \sqrt{t_{obs}}$. There are two ways to express the sensitivity, the integral sensitivity refers to the integrated flux above a selected minimum (threshold) energy, while the differential sensitivity addresses the flux in differential bins of energy. Figure 3.14 shows the sensitivity of MAGIC as of 2016.

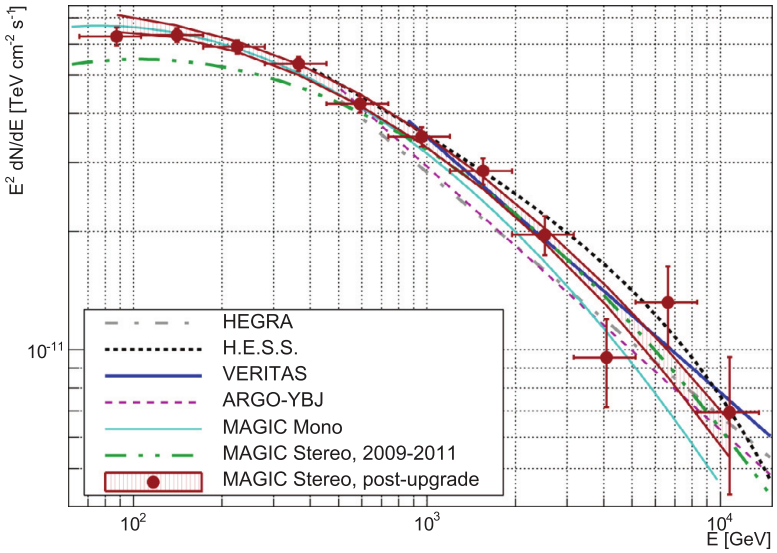


Figure 3.13: Spectra of the Crab Nebula measured with different instruments

Spectral energy distribution of the Crab Nebula obtained with the MAGIC telescopes after the upgrade (red points and shading) compared to other experiments: MAGIC-I (cyan solid), MAGIC Stereo 2009–2011 (green dot-dot-dashed), HEGRA (gray dot-dashed), VERITAS (blue thick solid), ARGO-YBJ (magenta, dashed) and H.E.S.S. (black dotted). The vertical error bars show statistical uncertainties, while the horizontal ones represent the energy binning. Image taken from [19].

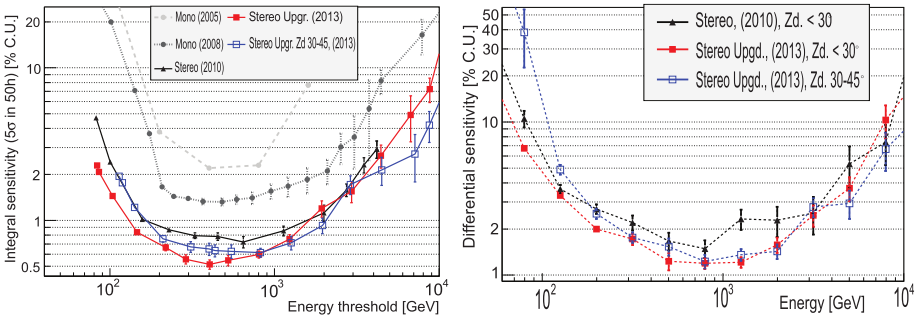


Figure 3.14: The sensitivity of MAGIC telescopes

MAGIC integral (left) and differential (right) sensitivity as a function of energy. The black lines are the sensitivities in 2010, before major upgrades in hardware, and the red and blue lines show the sensitivity after the last major upgrade in 2013. The color indicates the different pointing condition (red curves correspond to Zd observations below 30°, and the blue for the Zd in the range 30°-45°. Image taken from [19].

Chapter 4

Improvement of the energy reconstruction

This chapter describes the strategies I developed to improve the estimation of the energy of γ-rays in the data of the MAGIC telescopes.

First, the problem definition introduces an estimator, with the training process and important properties to be evaluated. Second, the Random Forest (RF) method, which plays the main role to improve the energy estimation, is introduced with the benefit to choose among a number of Machine Learning (ML) algorithms. There are different algorithmic approaches to RF in detail, and the approach I chose for the energy estimation will be highlighted, together with all the event reconstruction methods in MAGIC. The detailed explanation of the new energy estimator comes in the next, and the validation of the estimator will be followed. Finally, the performance evaluation shows the improvement of the energy estimation with respect to the standard strategy that has been used in the MAGIC collaboration during the last decade.

4.1 Problem definition

4.1.1 Definition of an estimator

Supervised learning

As mentioned in section 3.8 the event reconstruction focuses on the three event properties; the event class, the direction (Disp) and the energy. The determination of these γ-ray properties are done by the dedicated estimators, built by the dedicated algorithms.

Each algorithm builds an estimator as a mathematical model from a set of data that contains both the inputs and the desired outputs, and it will be applied to accurately generate outputs when given new inputs. This type is called supervised learning. Hereafter I discuss exclusively on a supervised learning problem.

Train data set

Let us consider a supervised learning task of estimating y from a set of input variables $\mathbf{x}$ on a sample in $\mathcal{D}$, where the pair of $(y, \mathbf{x})$ follows the joint probability function $y = f(\mathbf{x})$. However, there is uncertainty in the mapping from $\mathbf{x}$ to y, because of having limitation in the measurements of $(y, \mathbf{x})$ as well as not having enough attributes to sufficiently characterise the relation. Thus the relation needs additional term expressing the noise.

$$y = f(\mathbf{x}) + \varepsilon \tag{4.1.1}$$

where ε is noise which has zero mean and variance σ^2.

To construct the estimator, an algorithm learns the relation between y and $\mathbf{x}$ from a finite number of samples $\mathcal{D}_{train}$, called train sample. The train sample must "asymptotically" cover entire region of interest for an unseen sample in $\mathcal{D}$. Normally the region of interest corresponds to the probability distribution of $\mathcal{D}$. In that case the train sample $\mathcal{D}_{train}$ is chosen so that it is independent and identically distributed (i.i.d.) from $\mathcal{D}$.

After the estimator is constructed, it estimates y. The estimation is denoted with circumflex over the symbols in interest. The estimated value $\hat{y}$ for an event with a set of input variables $\mathbf{x}$ is the output of the model $\hat{f}$. Namely,

$$\hat{y} = \hat{f}(\mathbf{x}) \tag{4.1.2}$$

For notational convenience, hereafter abbreviation $f = f(\mathbf{x})$ and $\hat{f} = \hat{f}(\mathbf{x})$ will be used.

4.1.2 Training an estimator

Overfitting and generalization error

The training process is a fit of a model $\hat{f}$ to the train data. The complexity of the model is a measure of the ability of how precisely the model can adequately capture the underlying structure of the data. However, the accuracy of the estimation drops if the model is overly complex with respect to the noise in the data and the model captures the noise. This is called overfitting.

Therefore the performance of an estimator must be examined in terms of generalization error. It is a measure of how accurately an algorithm is able to predict outcome values for previously unseen data. And it is formulated as the absolute value or square of deviation of the estimation $\hat{y} = \hat{f}$ from the real value y, averaged over entire samples in $\mathcal{D}$. Here we define this as the expected Mean Squared Error (MSE).

$$MSE = \mathbb{E}\left[(y - \hat{f})^2\right] \tag{4.1.3}$$

Bias-variance decomposition

The effect of overfitting in training can be interpreted in the concept called the bias-variance decomposition [63, 84]. This concept shows that generalization

error is decomposed into two sides called "expected bias" and "model variance"[1].

For the decomposition I introduce $\mathbb{E}\left[\hat{f}\right]$, as the expectation of $\hat{f}$ over asymptotically infinite number of train samples to construct $\hat{f}$. As the train samples increases, the estimation $\hat{y}$ approaches y but the generalization error never goes to zero because the estimation learns from and is applied to the data with noise denoted in eq.(4.1.1). Because of the noise ε, the model can never know underlying function f in training, neither can match y which is randomly offset from f.

Using $\mathbb{E}\left[\hat{f}\right]$, MSE eq.(4.1.3) is decomposed as below (See section A for the detail).

$$MSE = \left(\mathbb{B}ias\left[\hat{f}\right]\right)^2 + \mathbb{V}ar\left[\hat{f}\right] + \sigma^2 \tag{4.1.4}$$

Where we define expected bias and model variance as below.

$$\mathbb{B}ias\left[\hat{f}\right] = \mathbb{E}\left[\hat{f}\right] - f \tag{4.1.5}$$

$$\mathbb{V}ar\left[\hat{f}\right] = \mathbb{E}\left[\left(\hat{f} - \mathbb{E}\left[\hat{f}\right]\right)^2\right] \tag{4.1.6}$$

Expected bias is the difference between the estimation by the expectation of model and the correct value which the model is trying to predict. The more complex the model, the closer the expectation of model to the correct function f and the model gets low expected bias.

Model variance is the "distance" of the model trained, from the expectation of model. Since the train samples are limited, the more complex the model, the more amount of train samples are needed to converge towards the expectation of model, thus there remains more variance. In other words, the less complex the model can make $\hat{f}$ already close to $\mathbb{E}\left[\hat{f}\right]$.

The generalization error has these two decomposed terms, and it shows which of correct function or expectation of model is closer to the model $\hat{f}$ constructed from the train data set.

Underfitting and overfitting

For a given amount of train data, the generalization usually develops with increase of model complexity as figure 4.1. The generalization error never approaches 0, instead it shows turnover at some point. This is called bias-variance tradeoff and related to the expected bias term and model variance term decomposed from the generalization error.

In the generalization error, a high expected bias indicates missing of the relevant relations between $\mathbf{x}$ and f in the estimation model $\hat{f}$ even after approximately infinite training. It happens when the model is oversimplified and does not capture the complexity of data. It always leads to high error naturally not

[1] Normally they are called just bias and variance. However, the names are very general and can confuse the reader with the other terms which can also be regarded as "bias" or "variance". The bias here in the focus of training process is different from the one mentioned later in the evaluation with test data set. I call the bias treated here as "expected bias" to discriminate. In the same way, I name "variance" here as "model variance".

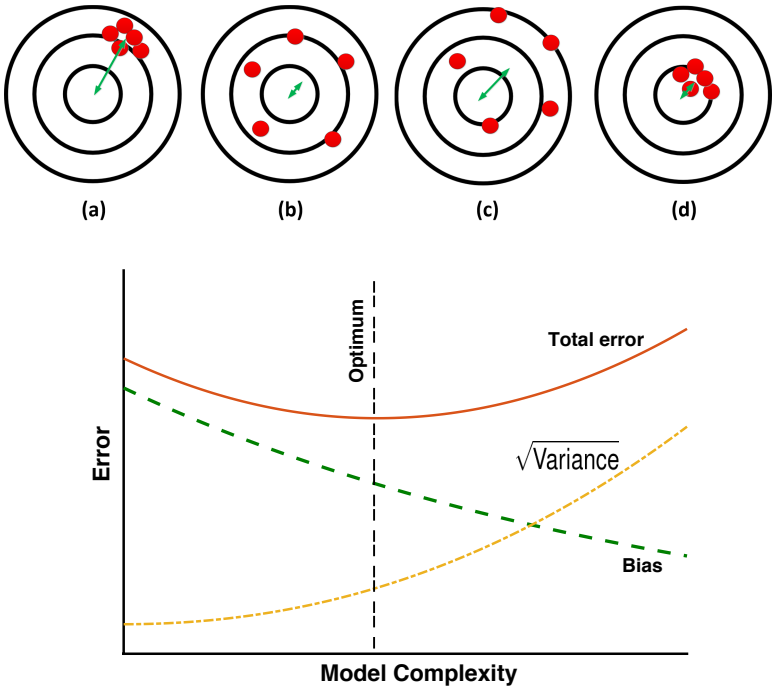


Figure 4.1: Bias-Variance tradeoff vs model complexity.

Top: The dart example for (a) high bias and low variance, (b) low bias and high variance, (c) high bias and high variance, and (d) low bias and low variance. The worst and best cases are (c) and (d), respectively. The center of the circles is the true value of the variable. Figure extracted from [57].
Bottom: The schematic of the bias-variance tradeoff. Bias, shown in the green dashed line, decreases as the model complexity increases. But the variance, yellow dashed line, increases instead. Thus the generalization error, the red solid line, upturns after an optimum point. The figure is redrawn based on [84]

only on training data but also on test data. This is called underfitting. Thus we prefer low expected bias; The model should learn as much in detail as possible the relation between the input and output for given train data set.

However, learning too much in detail results in capturing also the random noise of the training data, and the model cannot make the right prediction to unseen data. In this case, the model variance term enhances the contribution to the generalization error. If the model was too complex that can be tuned so flexible to the training set, it means the expectation of model is so close to the correct function (low expected bias) that it requires infinite amount of train data to converge the training. Therefore variance of the model $\hat{f}$ trained from the limited amount of train data is large with respect to the expectation of model. Consequently, a model with high model variance shows very small error to train data set, but large error to test data set. This is called overfitting. Thus we prefer low model variance; we also keep an eye on the variance so that both low expected bias and low variance are achieved to some fair extent.

The bias-variance trade-off is a central problem in supervised learning.

4.1.3 Evaluation of an estimator

Test data set

The evaluation of an estimator is ideally to use generalization error eq.(4.1.3), however it is impossible to collect all the unseen data. Instead it is approximated by another set of samples called test sample. Here I denote it $\mathcal{D}_{test}$. To properly predict the generalization error, the test data should be effectively independent and identically distributed from $\mathcal{D}$ [2].

$$MSE = \frac{1}{N} \sum_{X \in \mathcal{D}_{test}} (\hat{f} - y)^2 \tag{4.1.7}$$

Bias and resolution

For an evaluation of generalization error, two properties, bias and resolution will be more useful in this study. Because the output ranges widely, I express them in the form of normalised error.

Bias is the average difference between the prediction of our model and the correct value which we are trying to predict. Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.

$$Bias(\hat{f}, \mathcal{D}_{test}) = \frac{1}{N} \sum_{X \in \mathcal{D}_{test}} (\hat{f} - y)/y \tag{4.1.8}$$

Resolution, or standard deviation (SD), is the variability of model prediction for a given output value which tells us the spread of the model prediction. Namely, it is square root of variance as below.

$$\begin{aligned} Var(\hat{f}, \mathcal{D}_{test}) &= \frac{1}{N} \sum_{X \in \mathcal{D}_{test}} \left(\frac{\hat{f} - y}{y} - Bias(\hat{f}, \mathcal{D}_{test}) \right)^2 \\ SD(\hat{f}, \mathcal{D}_{test}) &= \sqrt{Var(\hat{f}, \mathcal{D}_{test})} \end{aligned} \tag{4.1.9}$$

Fitted bias and resolution

For more practical point of view, it is better to show bias and resolution in a form of representing the distribution of majority events. Especially in case of energy estimation, as described in section 3.10, the distortion of spectrum due to the energy estimation error needs to be corrected dependent on the migration matrix. In that case, the population becomes important measure.

An example of fitted bias and resolution is shown in figure 4.2. The normalised error distribution is fitted with a Gaussian. The bias and resolution are the peak and width extracted from the function. This works well for evaluating energy estimation, but this evaluation misses taking the "outlier" in the distribution into account, which is the case of previous method (LUTs-Erec). Thus for evaluating the outlier in distribution, SD will also be used.

[2] The selection of test data in the energy estimation is described in detail in section 3.4.3.

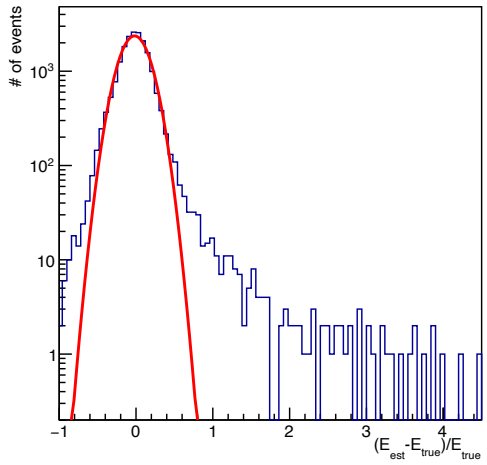


Figure 4.2: Bias and resolution

An example distribution of the normalised error,$(E_{est} - E_{true})/E_{true}$, for the energy estimation of γ-rays, in the energy range 1 TeV $\leq E_{true} \leq$ 2 TeV. The red curve is the Gaussian function fitted to the distribution. In this thesis, the peak position and the width of the Gaussian are referred to as "fitted bias" and "fitted resolution".

4.2 The advantage of Random Forest for the energy reconstruction among Machine Learning algorithms

Machine Learning (ML) is a method of statistical learning that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference of inputs instead. The learning process are classified into the three categories; supervised learning, unsupervised learning, and reinforcement learning. The study reported in this thesis focuses on the supervised learning.

Classification or regression More in detail, supervised learning are categorized to two algorithms; classification algorithms and regression algorithms.We adopt both algorithms depending on different purposes as below. Classification algorithms are used when the outputs are restricted to a limited set of values. This meets our problem of the event classification, that is to discriminate γ-ray-like events from the background events. On the other hand, regression algorithms are named for their continuous outputs, meaning they may have any value within a range. In case of the MAGIC analysis, the estimation of DISP for the direction reconstruction of γ-ray is categorized as this type of problem, and the energy estimation as well.

Choice of Random Forest There are a number of ML algorithms well accepted these days, however the selection of the optimal ML algorithm for a specific problem is not well established yet. There has been the attempts to compare the performances of the ML algorithms on classification problems [47, 91], and the results show that the performance of an algorithm is problem-dependent and there is no single guaranteed solution to outperform in all the problem cases. However, these studies show the Random Forest (RF) proves to be one of the algorithms with generally the best performance in a number of different problem settings.

RF has some advantages by its design [34], which was the reason to adopt it for the event reconstruction of VHE γ-ray observation in MAGIC. The merits can also be compared with the other methods [63]. Table 4.1 characterises the appropriate situations of each of the respective methods. In the comparison, the RF can refer to a "Trees" method, as it is an algorithm with ensemble learning approach to Decision Tree (DT) method. The "Trees" method also includes Boosted Decision Trees (BDTs) method, which performs the best in [91].

From the superior points of RF and the comparison with the other methods, the following is especially important to the event reconstruction.

- Versatile
  RF can be applied to both classification and regression problems.

- Robust to outliers in training samples
  RF is robust to outliers and non-linear data. The training of algorithm is not influenced by outliers and multicollinearity to some fair extent. Thus the training process requires less data cleaning and we can treat the simulation data with different conditions of the telescopes in fair way.

- Almost no risk of overfitting
  Training process of RF supervises the model to be with low expected bias, however at the same time it achieves moderate model variance by the randomization. Namely, RF is less likely to overfit on the training data.

- Easy to tune
  The performance of RF is already close to its best in general, with very little tuning required.

- Light and fast computation
  The size of the estimator is proportional to the number of trees. And RF requires less number of trees to obtain sufficient performance. RF tends to converge already at around 50 trees, which is significantly less than BDTs. Moreover the generation of RF can be parallelized, while BDTs need subsequent generation of trees. Therefore RF is lighter and faster.

- Interpretability
  Having Decision Tree method in the basis, we can take advantage of interpretability because the tree is structured with simple binary decisions. Moreover the training process tells the importance of each input variable. To be always watching the possible systematic effects or lack of statistics, having interpretability is very important.

Table 4.1: Some characteristics of different learning methods.

Characteristic	Neural Nets	SVM	Trees	MARS	k-NN, Kernels
Natural handling of data of "mixed" type	▼	▼	▲	▲	▼
Handling of missing values	▼	▼	▲	▲	▲
Robustness to outliers in input space	▼	▼	▲	▼	▲
Insensitive to monotone transformations of inputs	▼	▼	▲	▼	▼
Computational scalability (large N)	▼	▼	▲	▲	▼
Ability to deal with irrelevant inputs	▼	▼	▲	▲	▼
Ability to extract linear combinations of features	▲	▲	▼	▼	◆
Interpretability	▼	▼	◆	▲	▼
Predictive power	▲	▲	▼	◆	▲

Key:▲= good, ◆=fair, and ▼=poor. Learning methods: SVM=Support Vector Machine, MARS=Multivariate Adaptive Regression Splines, k-NN=k-Nearest-Neighbor Classifiers. The RF is in a family of Trees. The table is redrawn based on [63].

4.3 Random Forest

Random forest was first described by Breiman et al (2001) [34], and is now one of the most widely used and versatile algorithms in data science and machine learning [84]. It is based on Decision Tree (DT) method, and it overcomes bias-variance limit by the ensemble approach on a number of decision trees (DTs), which is the origin of "forest". And the ensemble approach features randomization strategy in building individual DTs, which is where "random" comes from. Thanks to this ensemble approach, RF achieves overwhelmingly more reliable prediction than DT. In bias-variance tradeoff point of view, training process of RF generates many models with low bias but large variance. The point is the models are randomized so that they are uncorrelated to each other. By combining the predictions, the large number of uncorrelated errors average out to zero to solve this problem

In this section, I introduce DT method first to be a basis for the RF method. Afterwards, the ensemble approach of RF to DT will be described, listing different actual approaches. At last, the RF already equipped in MAGIC analysis package ,"MAGIC Analysis and Reconstruction Software (MARS)", will be explained.

4.3.1 Decision Tree

A Decision Tree is an algorithm to make a decision on output, judging from a set of input variables via repetitions of different binary decisions. As the example (figure 4.3) shows, the entire schematic view of the process flow looks like a tree drawn upside down (See the description as well for the terminology). When an input is given, it traces the nodes which represent binary decisions, where one of the variables of the input is compared with the criterion value and is judged which side of the next node to be passed via the link (branch). When it reaches a terminal called a leaf, a single correspondent outcome is given as an output.

Construction and application of a tree

A DT is grown from a training samples $\mathcal{D}_{train}$ by the recursive processes of splitting them into two groups to construct new nodes. This process is summarized in algorithm 1. The application of the DT for the estimation is in algorithm 2.

The procedure starts from a set of training samples, which have the input variables paired with the desired output. The construction of the decision tree consists of the node building processes, each of which is done by splitting the samples into two groups, and the determinations of the predictive values at the end of splittings. The split focuses on one of the input variables. The samples are arranged in the sequence of the input variable and the split position is examined. The best split position should make the output values in each group the most different from each other. There are different methods to achieve this split and they will be described later. When a group of samples reaches the given condition to terminate further split, the terminal is defined with the prediction value decided from the samples in the group. In the followings, the detail of the termination condition and the prediction construction will also be separately discussed.

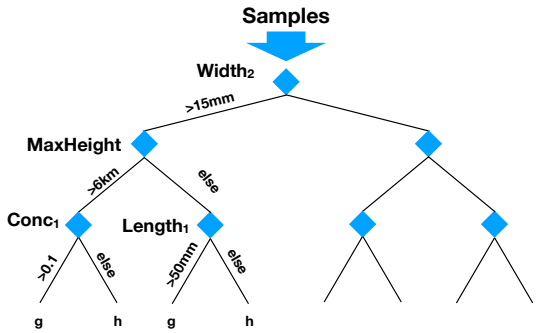


Figure 4.3: An example of Decision Tree

This example shows a sketch of a Decision Tree applied to the classification problem to separate γ-ray events (denoted as g) from the background events (denoted as h) in the observation data. A set of observation data is given at the top of the figure called "Root node", and will be evaluated by a set of questions. The questions are given one by one at the blue diamond points called "node", at each of which an attribute (variable) is compared with the criterion and passed to the correspondent branch. At the terminal node, the result of the evaluation (g or h) is given.

Algorithm 1 Construction of a Decision Tree

1: The set of training samples $\{X|X \in \mathcal{D}_{train}\}$ is given. An event has both the output and the p input variables paired. Namely, for an event,
$X = (y, \mathbf{x})$
$\mathbf{x} = (x^{(1)}, x^{(2)} \dots x^{(p)})$.
In the following steps, $\mathcal{D}_{train}$ will be split recursively.

2: Let n a node to generate a split and we define $\mathcal{D}_n \in \mathcal{D}_{train}$ to be a set of subsamples before a split and $\mathcal{D}_L, \mathcal{D}_R$ to be after, such that

$$\{\mathcal{D}_L \in \mathcal{D}_k, \mathcal{D}_R \in \mathcal{D}_k | \mathcal{D}_L \cap \mathcal{D}_R = \varnothing, \mathcal{D}_L \cup \mathcal{D}_R = \mathcal{D}_n\}$$

3: For splitting $\mathcal{D}_n$, a variable $x^{(j)}$ is chosen. Define two regions $\mathcal{R}_L$ and $\mathcal{R}_R$, as the left (lower) and right (bigger) sides of the split point s on $x^{(j)}$ over the samples.

$$\mathcal{R}_L(j, s) = \{X|X \in \mathcal{D}_n, x^{(j)} < s\}$$
$$\mathcal{R}_R(j, s) = \{X|X \in \mathcal{D}_n, x^{(j)} \geq s\}$$

Search for s to separate the y the best.

4: Step 3 is usually repeated for different variables, and the best $x^{(j)}$ and s will be chosen.

5: Evaluate if the recursive process reaches the condition to stop. If $\mathcal{R}_L(j, s)$ is still to be split again, $\mathcal{D}_n = \mathcal{R}_L(j, s)$ and repeat the procedure from step 2. The same rule is applied to $\mathcal{R}_R(j, s)$.

6: After $\mathcal{R}_L(j, s)$ or $\mathcal{R}_R(j, s)$ reaches the condition to stop the split, the node becomes a terminal (last node). A terminal t will be given a prediction value $\hat{y}_t$ from the population in t. When all the split processes reach terminal, the construction terminates.

Algorithm 2 Application of a Decision Tree

1: For an event in interest, the tree generated in algorithm 1 estimates $\hat{y}$, the value of the target variable y. The event has the same set of input variables as the training events $\mathbf{x} = (x^{(1)}, x^{(2)} \dots x^{(p)})$, and $\mathbf{x}$ is given to the tree.
2: The tree is traced branch to branch. On the track every node evaluate $\mathbf{x}$ and decide which side of branch to pass to. For a node n, the variable to evaluate and the split criteria $x^{(j)}$ and s are assigned in the construction of the tree. Using them, the event is traced to the side with the condition it satisfies.
3: The previous step is repeated until it reaches the last node. When the event reach the correspondent last node t for which $\hat{y}_t$ is assigned as the predictive value in the construction, the tree gives the predictive value, namely, $\hat{y} = \hat{y}_t$.

Split method for regression problems

The split method is dependent on if the problem is regression or classification. For a regression problem, minimization of weighted mean of variances with the events going to the two sides of node is used. The discussion below is based on the studies reported in the following publications [63, 69, 71].

Consider a train data set $\mathcal{D}_{train}$ with the number of events of N_{train}. The element $X \in \mathcal{D}_{train}$ has the p input variables $\mathbf{x}$ and desired output y. Namely, the i-th event can be denoted as below.

$$\begin{aligned} X_i &= (y_i, \mathbf{x}_i) \\ \mathbf{x}_i &= (x_i^{(1)}, x_i^{(2)} \dots x_i^{(p)}) \end{aligned} \tag{4.3.1}$$

To construct an estimator of y, the goal is to partition $\mathcal{D}_{train}$ into the segments with similar y. Assume that after the partitioning there are T non-overlapping regions.

$$\mathcal{R}_t \in \left\{ \mathcal{R}_1, \mathcal{R}_2, \dots \mathcal{R}_T \mid \begin{array}{l} \mathcal{R}_t \subset \mathcal{D}_{train} \\ \wedge \mathcal{R}_i \cap \mathcal{R}_j = \varnothing \forall i, j \in \{1, 2, \dots, T\} \\ \wedge \mathcal{R}_1 \cup \mathcal{R}_2 \cup \dots \mathcal{R}_T = \mathcal{D}_{train} \end{array} \right\}$$

Each $\mathcal{R}_t$ corresponds to a leaf, where the predictive value $\hat{y}_t$ is assigned. The predictive value needs to be chosen so that the Residual Sum of Squares (RSS) from output is minimized. So that,

$$\mathrm{RSS}_{\min}(\mathcal{D}_{train}, T) = \min_{\hat{y}_t} \sum_{t=1}^{T} \sum_{X_i \in \mathcal{R}_t} (y_i - \hat{y}_t)^2 \tag{4.3.2}$$

This means the fit with T values of constants to the distribution of y. Thus the best split search is an instance of the maximum likelihood approach. If we try to solve this by splitting all T regions, it is computationally infeasible. However, the problem is easy to solve when the regions to split is two and we can sufficiently partition $\mathcal{D}_{train}$ by recursive split processes.

For more in detail of the split, let us assume a node n to split the data $\mathcal{D}_n \in \mathcal{D}_{train}$. From the set of input variables $\mathbf{x}$ a variable $x^{(j)}$ is chosen. Define

the two regions $\mathcal{R}_L$ and $\mathcal{R}_R$, which are based on the split point s on $x^{(j)}$ over the samples.

$$\begin{aligned}\mathcal{R}_L(j,s) &= \{X|X \in \mathcal{D}_n, x^{(j)} < s\} \\ \mathcal{R}_R(j,s) &= \{X|X \in \mathcal{D}_n, x^{(j)} \geq s\}\end{aligned} \tag{4.3.3}$$

$$\begin{aligned}&\mathrm{RSS}_{\min}(\mathcal{D}_n) := \mathrm{RSS}_{\min}(\mathcal{D}_n, 2) \\ &= \min_{j,s}\left[\min_{c_L}\sum_{X_i\in\mathcal{R}_L(j,s)}(y_i - c_L)^2 + \min_{c_R}\sum_{X_i\in\mathcal{R}_R(j,s)}(y_i - c_R)^2\right]\end{aligned} \tag{4.3.4}$$

Where C_L and C_R are the solutions to minimize the square sum of residuals in left and right sides of the split. The minimum of C_L and C_R is achieved by the average of y in the respective side.

$$C_m = \hat{y}_m = \frac{1}{N_m}\sum_{X_i\in\mathcal{R}_m(j,s)} y_i \tag{4.3.5}$$

where N_m is the number of samples in $\mathcal{R}_m$.

So that,

$$\begin{aligned}\mathrm{RSS}_{\min}(\mathcal{D}_n) &= \min_{j,s}\left[\sum_{X\in\mathcal{R}_L(j,s)}(y_i - \hat{y}_L)^2 + \sum_{X\in\mathcal{R}_R(j,s)}(y_i - \hat{y}_R)^2\right] \\ &= \min_{j,s}\left[N_L\sum_{X_i\in\mathcal{R}_L(j,s)}\frac{(y_i - \hat{y}_L)^2}{N_L} + N_R\sum_{X_i\in\mathcal{R}_R(j,s)}\frac{(y_i - \hat{y}_R)^2}{N_R}\right]\end{aligned} \tag{4.3.6}$$

Thus the squared sum of error can be regarded as the weighted mean of population variances[3].

The reduction of the squared sum of error as below indicates the strength of the discrimination power for the regression.

$$\begin{aligned}&\Delta\mathrm{RSS}_{\min}(\mathcal{D}_n) \\ &= \sum_{i\in\mathcal{D}_n}(y_i - \hat{y}_n)^2 - \min_{j,s}\left[\sum_{X\in\mathcal{R}_L(j,s)}(y_i - \hat{y}_L)^2 + \sum_{X\in\mathcal{R}_R(j,s)}(y_i - \hat{y}_R)^2\right]\end{aligned} \tag{4.3.7}$$

The discrimination power for a variable j, or "importance of j", is defined as the sum of all the reductions for the nodes in the recursive split where j is selected.

[3] In the regression process of RF in MARS, sample variance is used and the denominator is $N_m - 1$.

Split method for classification problems

Unlike the regression problem to predict a quantitative output, the estimator for a classification problem predicts a qualitative output from a set of inputs in a form of a choice from a limited set of values. The predictive values should be the most commonly occurring class of observations in the region to which it belongs.

Thus training process for classification problem cannot define RSS and another measure for good split is needed. It can be defined by focusing on a proportion of a class observed in the node.

In a region $\mathcal{R}_m$, the proportion of class k among K classes is denoted as

$$p_{mk} = \frac{1}{N_m} \sum_{X_i \in \mathcal{R}_m} I(y_i = k)$$
$$\text{where } I(y_i = k) = \begin{cases} 1, & y_i = k \\ 0, & y_i \neq k \end{cases} \tag{4.3.8}$$

Instead of minimizing the variance, the target on the splitting is to make the separated regions to be more purely dominated by a class. In other words, we define a measure of impurity and minimize it. The simplest way to define impurity, called misclassification error, is to find the most dominant k and calculate the proportion of the rest as below.

$$\frac{1}{N_m} \sum_{X_i \in \mathcal{R}_m} I(y_i \neq k) \tag{4.3.9}$$

But this is not useful measure. Instead there are two of more amenable functions to apply, the Gini index and the cross-entropy. Compared to misclassification error, they are differentiable and more sensitive to changes in the node proportions (figure 4.4).

- Gini index (for classification)
  The Gini index is a measure of total variance of the population across all the defined classes. In a region $\mathcal{R}_m$ where K classes are defined, the Gini index [4] is

$$G_m = \sum_{k=1}^{K} p_{mk}(1 - p_{mk}) \tag{4.3.10}$$

  Gini index takes smaller value for the case with a more dominant class. For the split, Gini indexes in the both sides are calculated and the minimum of the average weighted by the population is searched. Namely, for the separation in the setting as eq.(4.3.3), the minimum averaged Gini index is

$$\begin{aligned} G_{\min}(D_n) &= \min_{j,s} \left[N_L G_L + N_R G_R \right] \\ &= \min_{j,s} \left[N_L \sum_{k=1}^{K} p_{Lk}(1 - p_{Lk}) + N_R \sum_{k=1}^{K} p_{Rk}(1 - p_{Rk}) \right] \end{aligned} \tag{4.3.11}$$

[4] In two-class case, eq.(4.3.9) can be converted a $p_1(1-p_1)+p_2(1-p_2) = 2p_1p_2 = 1-p_1^2-p_2^2$. Thus sometimes it is also written as $p_1^2 + p_2^2$.

Like eq.(4.3.7), "Gini decrease", the reduction of Gini, indicates the separation power.

- Information gain (for classification)
  Entropy of a distribution can be defined from information theory [5], as

$$S_m = -\sum_{k=1}^{K} p_{mk} \log_2 p_{mk} \tag{4.3.12}$$

The uniform distribution gives the highest entropy, and the more the distribution is skewed, the smaller the entropy. The purpose of split is to get the distribution with more purely dominated by a class, namely more skewed. Thus it can be interpreted as impurity. The impurity to be minimized is defined as weighted average similarly to the Gini index strategy.

$$\begin{aligned} S_{\min}(D_n) &= \min_{j,s} \left[N_L S_L + N_R S_R \right] \\ &= \min_{j,s} \left[N_L \sum_{k=1}^{K} p_{Lk}(-\log_2 p_{Lk}) + N_R \sum_{k=1}^{K} p_{Rk}(-\log_2 p_{Rk}) \right] \end{aligned} \tag{4.3.13}$$

The amount of entropy reduction after split can be the measure of better separation like eq.(4.3.7). This is called "Information gain".

Gini index and the cross-entropy are quite similar numerically and both of them are typically used to evaluate the quality of a particular split. In MARS, Gini index is used for the g/h separation. Gamma hadron separation tries to separate two classes, γ-ray (γ) and hadronic (h) events. Using the denotions of the propotions of them in the node m as below,

$$p_{m\gamma} = \frac{N_{m\gamma}}{N_m}, \qquad p_{mh} = \frac{N_{mh}}{N_m} \qquad \text{where } m = \{R, L\} \tag{4.3.14}$$

The weighted average of the Gini index is defined as

$$G = \frac{4}{N_L + N_R} \left(N_L p_{L\gamma} p_{Lh} + N_R p_{R\gamma} p_{Rh} \right) \tag{4.3.15}$$

Prediction value assignment

After the recursive split stops, the subsamples left in the node are used to make a prediction.

[5] In information theory eq.(4.3.12) is obtained in following way. "Information" is defined for an event to obtain, with the probability p of the event, as $-\log_2 p$. The lower the probability, the higher quantity the information the event carries. And the unit is called "bit". For the event with probability $p = 1/2$, information is 1 bit.
Entropy is the expected quantity of information to obtain, for an event x under probability distribution $p(x)$. Thus entropy S is,

$$\sum_x p(x)(-\log_2 p(x))$$

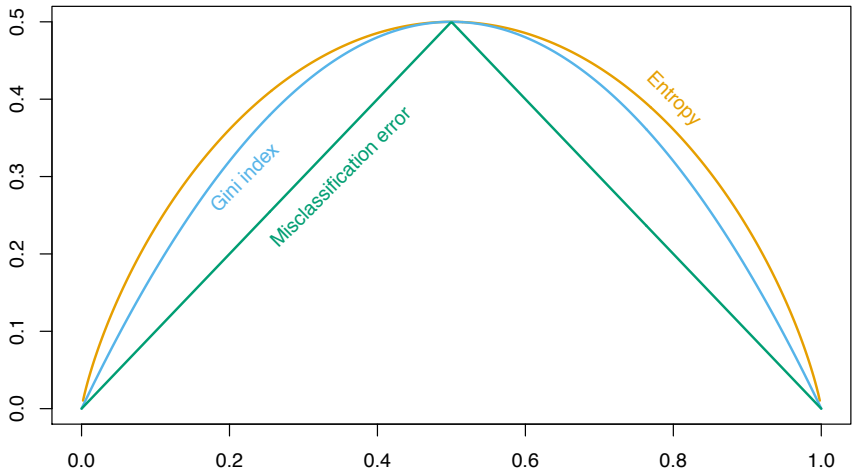


Figure 4.4: Node impurity measures in classification

Node impurity measures for two-class classification, as a function of the proportion p in class 2. Cross-entropy has been scaled to pass through (0.5, 0.5). Figure extracted from [63].

- Regression
  For a node $\mathcal{R}_t$, the predicted value $\hat{y}_t$ is assigned as

$$\hat{y}_t = \frac{1}{N_t} \sum_{X_i \in \mathcal{R}_t} y_i \tag{4.3.16}$$

- Classification
  In case of a classification problem, there are two ways. The simplest way is to assign the majority class in the node.

$$\hat{y}_t = \arg\max_k \sum_{i=1}^{N_t} I(y_i = k) \tag{4.3.17}$$

  It is also possible to assign a set of fractions in the population as the probabilities to the respective classes.

$$p(\hat{y}_t = k) = \frac{1}{N_t} \sum_{X_i \in \mathcal{R}_t} I(y_i = k) \tag{4.3.18}$$

- Classification of two classes
  In case the problem is binomial classification, the probabilities of the two classes are conjugate thus assigning y as following,

$$y = \begin{cases} 1, & \text{for class 1} \\ 0, & \text{for class 2} \end{cases} \tag{4.3.19}$$

  and calculating $\hat{y}_t$ as eq.(4.3.16), it is possible to give a single value as the probability to the class.

Regularization to avoid overfitting

It is clear that more complex decision trees lead to finer partitions and give better performance on the training set. In theory, one can grow the tree as deep as possible until the number of last nodes becomes the same as the number of training samples. However, as discussed in the bias-variance trade-off, this generally leads to overfitting and show larger generalization error on the test set. In other words, the algorithm becomes "noisy", causing it to model the random noise in the training data. Therefore decision trees need some form of regularization to control complexity and reduce overfitting. The regularization can be done in two ways.

- Setting constraints on growing a tree
  If growing tree is stopped well before the tree becomes too deep, the predictive model can avoid overfitting. There are many ways to stop the growing as below.

  1. Minimum node size (before split)
     Defines the minimum number of samples in a node. It can be used to judge if the growing will continue by the next split. Higher values prevent a model from learning detailed relations which might be highly specific to the particular sample selected for a tree. On the other hand, too high values can lead to under-fitting.
  2. Minimum node size (after split)
     If there is too few samples in one side after split, the samples can be noise. The split would be canceled and the node becomes terminal.
  3. Maximum depth of tree (vertical depth)
     The maximum number of the iteration of split. It can also be expressed in the form of "Maximum number of terminal nodes", as the number of terminal nodes is 2 to the power of the depth.
  4. Number of trials for a split
     The number of variables to consider while searching for a best split. In a split, the variables will be randomly selected under the limitation of iteration on this number, and the best split will be chosen. In this way, the split can avoid being too sensitive to a variable in the training samples. This strategy is one of the features of Random Forest.

- Tree Pruning
  This strategy first let the tree grown to a large depth. If the model is evaluated with the training data set itself then, the model with no expected bias and the prediction error would be 0. However, it has high variance and the prediction error becomes very large if it is evaluated with another data set. Then the proper depth is searched backwards, starting at the bottom and removing the nodes one by one until the error becomes minimum.

4.3.2 Ensemble approach of RF

As mentioned in the previous section, DT has a strong risk of overfitting and it cannot be free from the effort to balance the bias-variance trade-off by itself. However it can be overcome via an ensemble approach. An ensemble approach is to combine predictions from many predictive models. The models are generated rather deeply so that they become noisy but approximately unbiased models (They are also called weak learner). When the predictions by individual predictive models are combined, the noise will be canceled out and the prediction obtains low expected bias and low variance. To sufficiently decrease the noise from individual predictive models, the critical concept to ensemble approach is to have many uncorrelated models. And the ensemble approach via RF is one of the most successful strategies.

In case of Random Forest, the key to de-correlate the individual predictive models (trees) is to give two forms of randomization in the generation. One is the randomization in the samples to feed for the training of a tree, called "Bagging (Bootstrap Aggregating)". And the other is the randomization in the variable selection in the node splitting. The summaries of the algorithms of the training process and application process are described in algorithm 3 and algorithm 4 respectively, and the details of these strategies are described is the below.

Algorithm 3 Construction of a Random Forest

1: The set of training samples $X \in \mathcal{D}_{train}$ is given to the generation process
2: For a tree b, a set of subsamples $\mathcal{D}_b$ is generated from $\mathcal{D}_{train}$ by the Bagging process.
3: The tree b is generated from $\mathcal{D}_b$ via the same process as DT described in algorithm 1, but with random variable selection. For the generation of a branch, the limited number of random choice of split variable is performed. The best split position of the samples is searched for each variable. The variable with best split power is chosen for the branch generation.
4: Step 2 and 3 is repeated for B times.

Algorithm 4 Application of a Random Forest

1: For an event in interest, the RF generated in algorithm 3 estimates $\hat{y}$, the value of the target variable y. The event has the same set of input variables as the training events $\mathbf{x} = (x^{(1)}, x^{(2)} \dots x^{(p)})$, and $\mathbf{x}$ is given to the RF.
2: In the RF, $\mathbf{x}$ is given to all the trees. A tree b predicts $\hat{y}_b$ from $\mathbf{x}$. The process is described in algorithm 2
3: An ensemble is performed to combine predictions $\{\hat{y}_b | b = 1, 2, \dots B\}$ and the estimation $\hat{y}$ is obtained.

Bagging (Bootstrap Aggregating)

The ensemble method ideally requires the use of independent samples for generating individual trees. But it means the train data set needs to be divided to feed to each tree generation, and each tree cannot gain low expected bias

due to lack of statistics (with very less amount of data, a model is unable to capture the underlying pattern of the data). However, the bagging approach to the train data set gives a way to emulate the process of obtaining approximately new distinct sample sets keeping statistics.

BAGGing, or Bootstrap AGGregatING, first introduced by Leo Breiman [33], is a sampling technique in which one randomly samples with "replacement" from the data set. More specifically, from the original train set $\mathcal{Z} = \{X_1, X_2, \ldots X_N\}$, a new set of data set $\mathcal{Z}^{*1} = \{X_1^{*1}, X_2^{*1}, \ldots X_N^{*1}\}$ is generated picking the same number of samples at random, which possibly contains repetitive elements (figure 4.5).

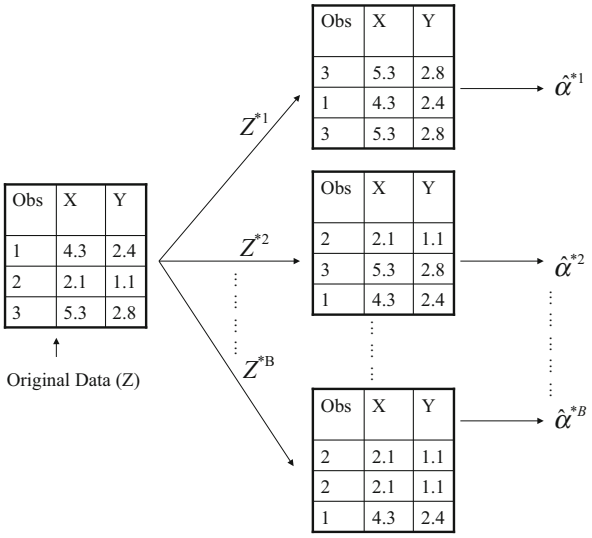


Figure 4.5: An example of bootstrap sampling

A graphical illustration of the bootstrap approach on a small sample containing $N = 3$ events. There are B bootstrap samples generated, each of which also contains N events, sampled with replacement from the original data set. Each bootstrap data set is used to obtain an estimate of target variable $\hat{\alpha}^{*b}(X)$. Image is taken from [71].

Because of the duplication, there will be samples that do not appear in the bootstrap samples. Consider there is N samples in train set and a set of subsamples will be generated by the bootstrap procedure. After the trials of random sampling for N times allowing a sample to be chosen multiple times, the probability of being NOT selected is $(1 - 1/N)^N$ and it asymptotically approaches $1/e$. This means, only $1 - 1/e \sim 63\%$ of N total number of samples are selected for a large number of the samples, and the rest are duplicate samples.

The samples which were not selected are called "Out-Of-Bag (OOB)" samples. These samples are independent from the bagged samples thus can be used for cross validation for the tree trained with the bagged samples.

Random variable selection

When the trees are trained using different bootstrap samples, the correlation of the trees is weakened (it is called a bagging trees). However, the bagging trees are still strongly correlated to each other because the bagged samples are not enough independent from each other, and they cannot have a significant improvement from using one tree [57]. To overcome this problem, Random forest addresses this issue by the random variable selection [43].

The random variable selection is the strategy for the split process in the training. In a normal decision tree, when it is time to split a node, every possible variables are considered and the one that produces the best separation will be selected. Therefore the models will be grown as similar splits. On the contrary, the random variable selection is the strategy to select limited number of variables at random at each node for the split, and choose the best split among them. Limited by the number of trials of random variable selection, the trees will be forced to grow as more variation in the models among the trees. In the end, it ultimately results in lower correlation across trees and more diversification and thus results in a noticeable improvement.

The number of trials of random variable selection affects the performance. The best number of trials m_{sp} needs to be much smaller than the number of variables p. Typically $m_{sp} \sim \sqrt{p}$ is chosen [71]. But the optimal range of m_{sp} is usually quite wide [35].

Combining predictions

The trees are trained in the same randomization strategy, originally from the same train set. The expectation of an average of B such trees is the same as the expectation of any one of them. This means the expected bias of bagged trees is the same as that of the individual trees, however the averaging improves the estimation error through variance reduction. The actual averaging strategy is as follows.

From a train set $\mathcal{Z}$, imagine we take B sets of subsamples $\{\mathcal{Z}^{*1}, \mathcal{Z}^{*2} \dots \mathcal{Z}^{*B}\}$ via bagging to construct a forest consist of B trees, which can be regarded a set of functions $\{\hat{\alpha}^{*1}, \hat{\alpha}^{*2} \dots \hat{\alpha}^{*B}\}$. A b-th tree predicts an output variable y from a set of inputs $\mathbf{x}$,

$$\hat{y}_b = \hat{\alpha}^{*b}(\mathbf{x}) \tag{4.3.20}$$

From the outputs from the B trees $\{\hat{y}_b\}$, the combined prediction $\hat{y}$ is calculated. There are mainly two strategies to combine these B predictions.

- Majority vote
  The forest chooses that classification having the most out of N votes.

$$\hat{y} = \arg\max_k \frac{1}{B} \sum_{b=1}^{B} I(\hat{\alpha}^{*b}(\mathbf{x}) = k) \tag{4.3.21}$$

  This is possible for only classification problem to predict single choice of class from discrete set of classes.

- Average
  The forest just the averages all the individual predictors.

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} \hat{\alpha}^{*b}(\mathbf{x}) \tag{4.3.22}$$

  This is suitable for continuous predictors like regression. But it is also applicable to classification if there are only two classes, as it can work as binomial probability.

- Weighted average
  It is also possible to introduce weight in averaging the predictions from the individual trees. In the generation of a tree, the node split does not finish with equal number of samples in all the last nodes, as long as minimum node size is not 1. The weight is set using the population at the last node when the tree is generated.

RMS of predictions

In case the prediction is performed as an average of the predictions, RMS of prediction can be defined.

$$\sigma_{\hat{y}} = \frac{1}{B}\sqrt{\sum_{b=1}^{B}(\hat{y}_b - \hat{y})^2} \tag{4.3.23}$$

This is not a measure of uncertainty. The distribution of prediction error does not have a reasonable correlation with RMS of prediction. The estimation values by individual trees varies primarily due to the randomization in training, therefore the RMS of predictions would indicate how well the trees are decorrelated. This means that the RMS of predictions can be used for a pathological purpose. One of such possibilities is an evaluation of statistical coverage on training data. The detail is discussed in section E.1.

Forest-growing parameters

Unlike DT, RF has almost no risk of overfitting, consequently, there is almost no strict regularization needed. However, there are some parameters available for tuning the performance. Here I name these parameters "forest-growing parameters".

- Number of trees
- Number of trials
  The number of random selection of variables for a split step.
- Minimum node size
  Number of events at last node. This number sets the condition to stop splitting a node.

4.3.3 RF in MAGIC

The RF reconstruction was implemented inside the MARS analysis package for MAGIC collaboration originally by Thomas Hengesthebek [65], and reported in [13]. The functionality is developed as C++ adoption of the original Fortran code by Leo Breiman and Adele Cluster [34].

As introduced in section 3.8, there are three event properties to reconstruct in the standard MAGIC analysis; particle-species class, incoming direction, and energy. The first two are determined with RF, and they work as below.

Direction through the parameter Disp

The parameter "Disp" is estimated separately for the individual telescope images. For the Disp of telescope $k(k = \{1, 2\})$,

- Input data
  MC train data is used.
- Target value
  $TrueDisp_{\text{tel.k}}$
- Input variable
  - $\log_{10}(Size_{\text{tel.1}})$
  - $\log_{10}(Size_{\text{tel.2}})$
  - $Width_{\text{tel.k}}$
  - $Length_{\text{tel.k}}$
  - $ClassicalImpact_{\text{tel.1}}$
  - $ClassicalImpact_{\text{tel.2}}$
  - $ClassicalMaxHeight$
  - $|TimeGradient_{\text{tel.1}}|$
  - $|TimeGradient_{\text{tel.2}}|$
  - Zd
- Event selection
  For the events which succeed in stereo reconstruction,
  $50 \leq Size_{\text{tel.1}} \leq 50000$ & $50 \leq Size_{\text{tel.2}} \leq 50000$
  & $Leakage1_{\text{tel.1}} < 0.15$ & $Leakage1_{\text{tel.2}} < 0.15$
- Node split criteria
  Weighted average of the variances
- Prediction assignment of a tree
  Average of the target values in the last node.
- Ensemble of the predictions
  Average of the predictions by all the trees.
- Forest growing parameters
  - Number of trees : 100
  - Number of trials : 3
  - Minimum node size : 5

Particle-species classification through the hadronnes parameter

- Input data
  MC train data is used for γ-ray samples and observation data is used for hadron samples. The characteristics of a shower image are dependent on the pointing direction and Size. To make the RF able to discriminate in different pointing direction and Size, the distribution of input data is adjusted so that the number of events are the same between the γ-ray samples and hadron samples in each bin in the three-dimensional histogram, with the dimensions and binning below:
  - $\log_{10}((Size_{\mathrm{tel.1}} + Size_{\mathrm{tel.2}})/2)$: 90 bins from 1 to 5.5
  - cos(Zd) : N_{Zd} bins from 0 to 1
  - Az : N_{Az} bins from 0 to 2π

  N_{Zd} and N_{Az} are configurable with default values of 30 and 1 respectively.
- Target value
  γ-ray samples as 0 and hadron samples as 1.
- Input variable
  - $\log_{10}(Size_{\mathrm{tel.1}})$, discretized in an interval of 0.05
  - $\log_{10}(Size_{\mathrm{tel.2}})$, discretized in an interval of 0.05
  - $Width_{\mathrm{tel.1}}$
  - $Width_{\mathrm{tel.2}}$
  - $Length_{\mathrm{tel.1}}$
  - $Length_{\mathrm{tel.2}}$
  - $ClassicalImpact_{\mathrm{tel.1}}$
  - $ClassicalImpact_{\mathrm{tel.2}}$
  - $|TimeGradient_{\mathrm{tel.1}}|$
  - $|TimeGradient_{\mathrm{tel.2}}|$
  - cos(Zd) discretized into 30 levels with equal intervals
- Event selection
  For the events which succeed in stereo reconstruction,
  $50 \leq Size_{\mathrm{tel.1}} \leq 50000$ & $50 \leq Size_{\mathrm{tel.2}} \leq 50000$
  & $Leakage1_{\mathrm{tel.1}} < 0.15$ & $Leakage1_{\mathrm{tel.2}} < 0.15$
  & $NumIslands_{\mathrm{tel.1}} = 1$ & $NumIslands_{\mathrm{tel.2}} = 1$
- Node split criteria
  Gini index.
- Prediction assignment of a tree
  Average of the target values in the last node.
- Ensemble of the predictions
  Average of the predictions by all the trees.
- Forest growing parameters

- Number of trees : 100
- Number of trials : 3
- Minimum node size : 5

4.4 The new methodology with Random Forest for the energy reconstruction

In this section I introduce the novel strategy of energy estimation using Random Forest (RF-Erec). First, the selection of input variables is described in section 4.4.1. The selection follows the insight into the nature of EAS and its geometrical relation with the telescopes, so that the energy reconstruction becomes robust and transparent to possible systematic effects. In the section 4.4.2 and section 4.4.3, the treatment of the input and output variables in the estimator is explained. For better performance, some variables need to be considered about how to be treated in the estimator. Section 4.4.5 summarizes the training and application processes. Among them are the flow of event reconstruction procedure, the summary of input variables and their importance in different Zd ranges. I also introduce the improvement of Disp in section 4.4.6, that of RF-Erec also contributes to the improvement of energy estimation. All these above are summarized in section 4.4.7, where the description about intermediate versions is included. At last, the implementation in the MAGIC standard analysis software is described in section 4.4.8.

4.4.1 The selection of input variables

The basis for the energy estimator is to define the input variables and the output. The output is of course the energy, and the input variables are all the parameters of an event that are correlated to the energy. To this end, RF has a great advantage since its training process gives the importance of each input variable, as well as it has good dimensionality of input variables and robustness to outlier samples. However, it is impossible to point a missing important variable, thus the investigation of the input variables is mandatory. Moreover, all the input variables need to be well controlled so that the risk of a systematic effect can be minimized, especially for the following points.

- Since it is impossible to artificially produce real VHE γ-ray event for calibrating the telescopes, the supervision data for the energy estimators must be produced by Monte Carlo simulation (MC). However it cannot perfectly reproduce real γ-ray event and it is inevitable to have discrepancies to the actual observation data. Therefore it is necessary to choose only relevant variables under physical view, otherwise some variables may be systematically different from reality and would bias the estimation.

- The variables chosen for the energy reconstruction need to be tied together under the physical view to minimize the systematic effects. They can arise from various reasons like statistical coverage, limited precision and so on. The physical view can help with identifying the sources of the systematic errors. In this thesis, some identification of systematic effects are done in this way.

Therefore, in this study, I introduce the integrated view of the input variables on a shower event not to miss the relevant variables, as well as to limit the amount of input variables not to introduce unknown effects.

Here, I remind the basic geometrical view of a shower to make sure to select all the important variables to estimate the energy of the γ-ray. The list of variables chosen are similar to the LUTs-Erec, and the view of the relation between the energy and the variables are also similar. But reviewing the geometrical view of the shower leads to additional variables and a better usage of them.

Size to energy and important corrections

As discussed in section 3.8.3, the most dominant variable to determine the energy of γ-ray is the total light content, which is primarily indicated by the parameter "Size". Size has already very strong linear correlation to the energy, however there are two obvious ideas on how they should be corrected for better correlation.

Firstly it is obvious that a fraction of events suffer from the limited field of view. The Size cannot explain the total light content if the image is truncated at the edge of the field of view. Thus the Leakage variables are indispensable to correct the truncated images.

Secondly, the size-to-energy relation needs further corrections based on the distance from the telescope to the location of the shower. Owing to the directionality of the Cherenkov light with the opening angle of $\sim 1^\circ$, the total amount of photons will be preserved roughly in the disk-like region. However the opening angle makes the density enhancement at the outer region of the disk. Thus the size-to-energy relation needs correction using lateral offset, i.e. how much the shower axis is lateral offset from a telescope (Impact). Moreover, the opening angle causes linearity between the emission height and illuminated region, the photon density to illuminate changes dependent on emission height of the photons. Thus the size-to-energy relation needs correction using the information of longitudinal offset, i.e. how much the shower is offset along the shower axis (MaxHeight).

The lateral distribution of the photon density is shown in figure 4.6. Although the Cherenkov density is constant at the close distance from the Impact point, the tail of the radial distribution follows roughly inverse proportion, as the sum of each particle's radiations is still reflecting the conical radiation. Therefore the distribution can be expressed roughly in a form of a function below,

$$d(x) = \frac{1}{2}\left(1 - \tanh(x - x_c) + \frac{1}{(x/x_c)}(\tanh(x - x_c) + 1)\right) \tag{4.4.1}$$

where x is Impact [m] and x_c is the lightpool radius.

This function of Impact can be used as the correction factor to Size, denoted as below,

$$Energy = C \cdot Size/d(x) \tag{4.4.2}$$

where C is a constant.

It already works well as the correction factor to the events with different Impact, and the corrected size obtain stronger linearity to energy as shown in figure 4.7. Therefore, it is advantageous to use the quantity Impact in the energy estimator.

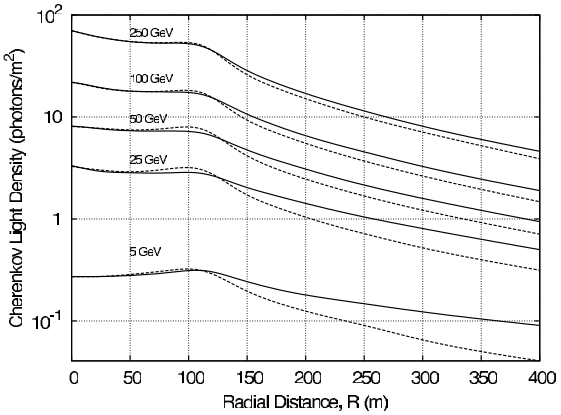


Figure 4.6: Cherenkov light density vs radial distance (Impact)

Actual simulations of the lateral distribution of Cherenkov light as a function of the radial distance from the shower axis (Impact). The dashed lines are the case when shower axis is along the Earth's magnetic field direction and the solid lines are perpendicular. Simulations were done using the Altai code assuming an observatory elevation of 2600 m. Figure extracted from [10].

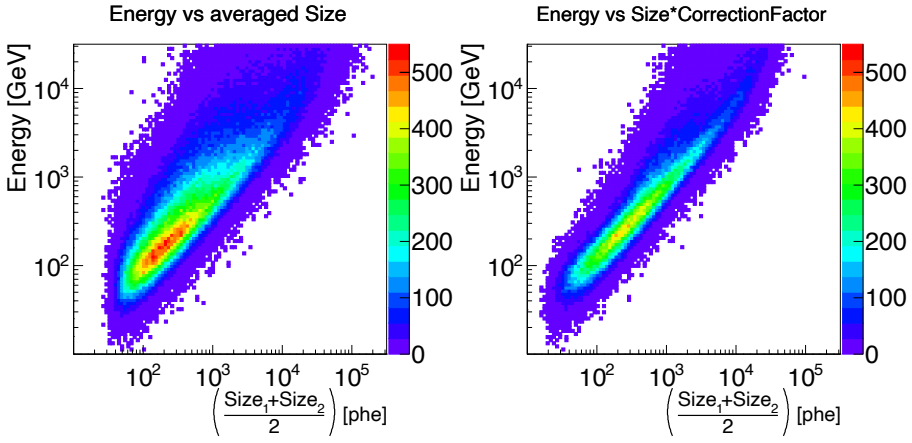


Figure 4.7: Energy vs Size with and without Size correction

The two-dimensional histograms of simulated γ-rays, displaying true energy vs. averaged Size, before (left) and after (right) the Size correction, mentioned in eq.(4.4.2), is applied.

On the other hand, the longitudinal offset (MaxHeight) varies shower to shower, dependent not only on energy but also on the first interaction point which is governed by the probabilistic process, although the expected height of first interaction is almost constant as shown in figure 2.3. The higher the MaxHeight, the more distant the brightest location in the shower, thus the dimmer the image (smaller Size). The relation is seen in the simulated data, shown in figure 4.8. Therefore it is advantageous to use the MaxHeight in the energy estimator.

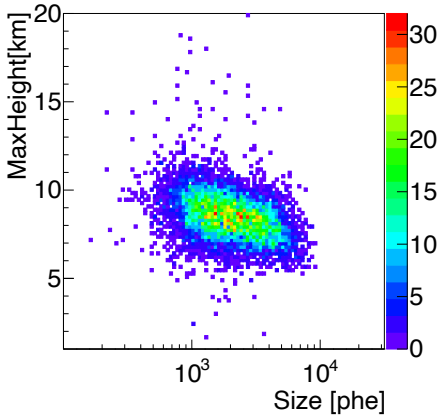


Figure 4.8: MaxHeight vs Size

The simulated γ-ray events in a two dimensional histogram to show the relation between MaxHeight and Size. In this example, the events selected are with true energy of 1-2 TeV, true Impact of 100-200 m for both telescopes.

From the inspection of the effects of the Impact and the MaxHeight above, it is clear that they provide useful information to obtain better energy estimation. And the reconstruction process has the estimated Impact and MaxHeight, via the geometrical reconstruction of three-dimensional shower profile. However their precision is limited and even becomes insufficient in high Zenith distance, thus we must collect the related variables to indicate them. Moreover their correlation to Impact or MaxHeight change mainly dependent on the energy range and the Zenith distance range. In the following sections, firstly the precision of Impact and MaxHeight is discussed and the related variables are introduced later, discussing also on the precision and the dependencies.

Impact and MaxHeight

Impact and MaxHeight needs to be discussed further. As discussed in section 3.7.2, Impact and MaxHeight are calculated in a integrated way as geometrical reconstruction of three-dimensional shower profile. The calculation is based on the reconstructed incoming direction, and there are two ways of direction reconstruction; classical method and Disp method. In general, the Impact calculated by Disp method (DispImpact) is much more accurate than that by classical method (ClassicalImpact). However the accuracy is dependent on the conditions and the possible systematic effect needs to be noted.

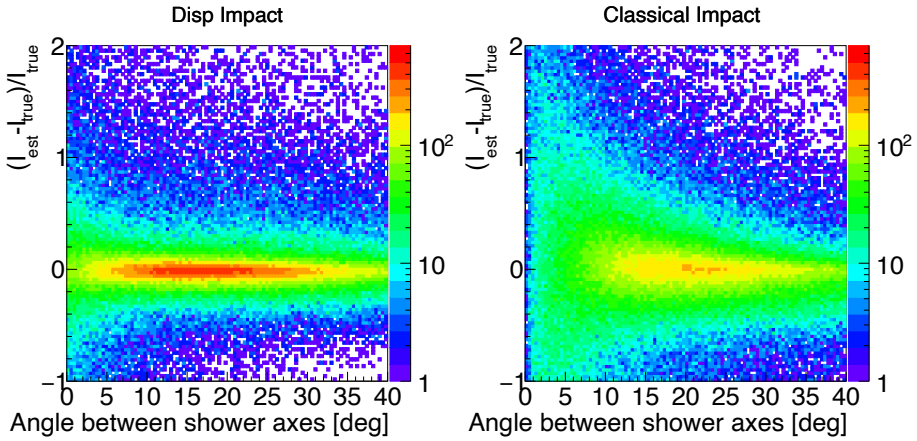


Figure 4.9: Accuracy in determination of DispImpact and ClassicalImpact

Two-dimensional histogram of γ-rays by MC simulation, showing normalized Impact estimation error, $(I_{est} - I_{true})/I_{true}$, where I_{est} is the estimated Impact and I_{true} is the true Impact, as a function of angle between shower axes, which is the crossing angle of shower image axes in superposed field of views of two telescopes. The left figure is for the estimated Impact via Disp method (DispImpact) and the right is via Classical method (ClassicalImpact).

Figure 4.9 shows the comparison of the accuracy of DispImpact and ClassicalImpact, presented as a function of the crossing angle of the axes of two shower images, as it is the most important dependence to affect direction reconstruction. As the Disp method improves the reconstruction of shower incoming direction, especially when the crossing angle is small, the accuracy comparison of Impact estimations also shows the most dramatic improvement in the small crossing angle.

Thus there is no doubt to choose DispImpact instead of ClassicalImpact, and it had been already adopted to LUTs-Erec. However there is always limited accuracy and it introduces systematic effects, thus it needs to be treated carefully. The systematic effects from the treatment will be discussed in section E.3.1 and the treatment in the new energy estimator will be discussed in detail in section 4.4.2.

The accuracy of DispImpact as the dependence on the TrueImpact is shown in figure 4.10. Out of the four figures, the top left shows the overall accuracy. The performance degrades in small Impact, although it should not heavily affect the energy estimation performance because the Cherenkov density is relatively flat inside the light pool. The other three figures are the same histograms but filtered in the energy of the γ-ray, 0.1-0.2 TeV, 1-2 TeV and 10-20 TeV. We can see that the accuracy has energy dependence worsening in the low energy. This would affect the estimation performance in the lowest energy range.

The accuracy is dependent also on the zenith distance. For the large zenith distance, the determination of the incoming direction cannot give enough precision for the determination of geometrical information especially on Impact, because the shower develops at very large distance from the ground, and the small fluctuation of the shower axis results in the large fluctuation of impact

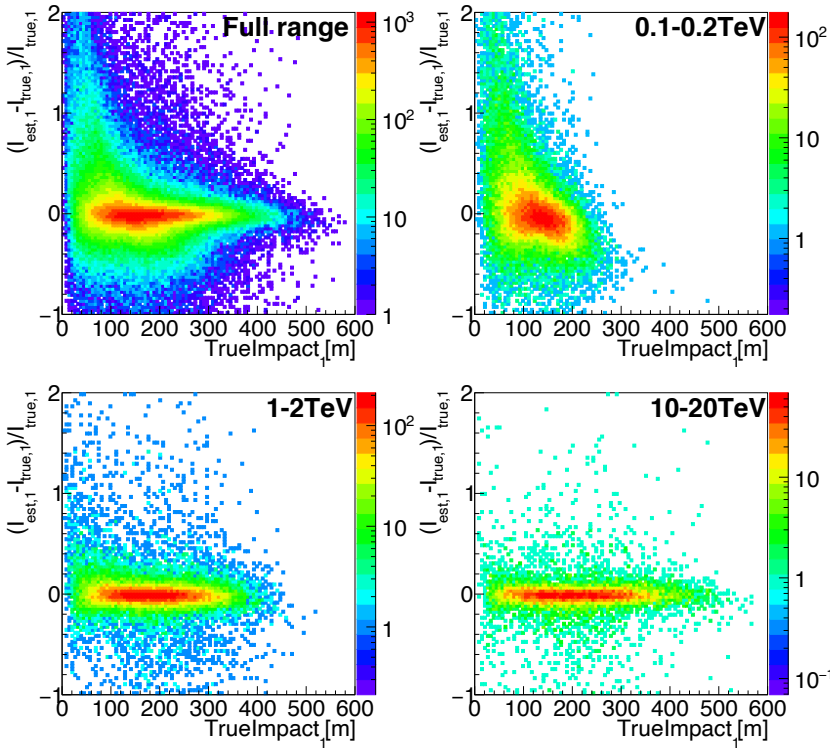


Figure 4.10: Normalised Impact error vs Impact

The accuracy of Impact estimation as a function of Impact is shown by normalized Impact estimation error vs TrueImpact of M1, for simulated γ-rays of different energies.

point at the order of several hundreds of meters in worst case of zenith degree over 70°. This effect is discussed in detail in section E.3.1.

Therefore, there needs to be alternative variables to support the poor Impact estimation especially in the small Impact range, low energy range and high zenith distance range.

As for MaxHeight, the calculation in the geometrical reconstruction is more indirect than Impact because the process is based on the Impact point as well as the reconstructed incoming direction, Thus the precision is obviously lower than that for Impact. Here I do not include the investigation on the accuracy of MaxHeight. In the next section, the correlated variables with Impact and MaxHeight will be discussed.

The variables to support Impact and MaxHeight

The variables listed here are correlated with Impact and/or MaxHeight, thus help the energy estimation with the geometrical correction to Size.

- Disp
  Disp is strongly correlated to both Impact and MaxHeight because it is the origin of the direction reconstruction, and hence the geometrical

reconstruction. The relations between Disp and Impact or MaxHeight can be understood intuitively in the following way. Disp should be larger either when the shower axis is more distant laterally (large Impact) as figure 4.11, or when the shower maximum location is lower longitudinally along the shower axis (low MaxHeight) as figure 4.12. Both relations can be seen clearly in the figures.

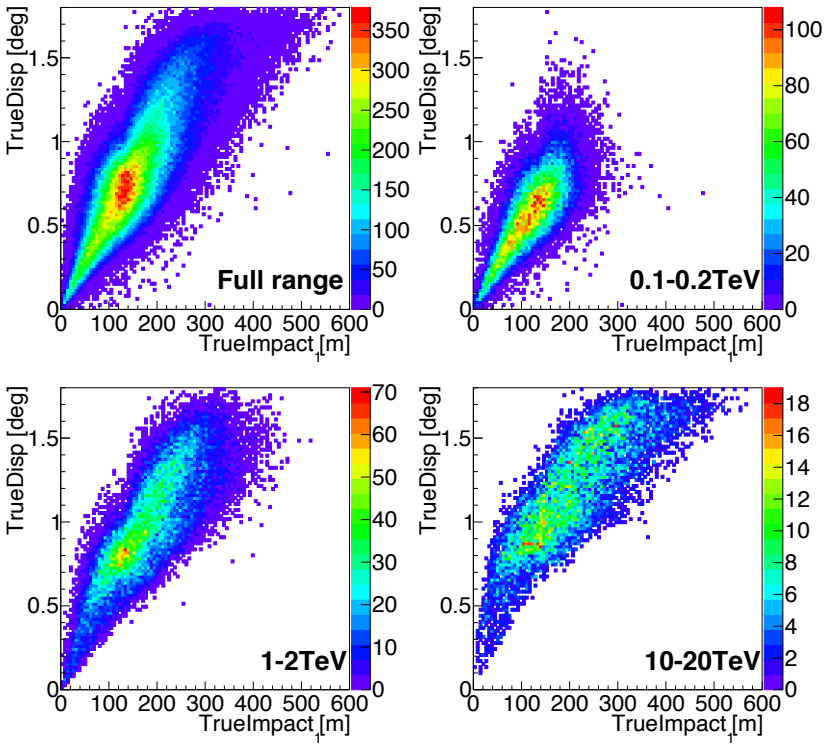


Figure 4.11: Correlation of TrueDisp with TrueImpact

TrueDisp vs TrueImpact of M1 for simulated γ-rays of different energies.

The accuracy of Disp is shown in figure 4.13. It degrades in the edges of the distribution range, where the distribution of estimated value is suppressed. In addition, the precision in low energy is worse. To deal with these issues, the variables correlated with Disp should be included. Since Disp is estimated using RF the input variables to the Disp estimator. The Disp is estimated mainly by Length, Width and TimeGradient. Of course they are also correlated to Impact and MaxHeight. They are described in the following.

- Length and Width
  As a characterization of the image shape, Length and Width should work together as a measure of the elongation of the image. Intuitively the image should be more elongated either when the lateral offset is large (large Impact) or the shower develops deeper (low MaxHeight).

  As the true value is available in Impact, I show the correlation the

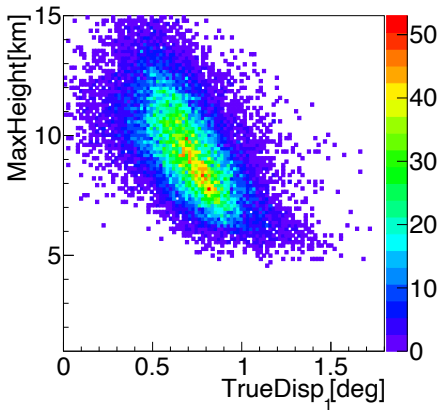


Figure 4.12: Correlation of TrueDisp with MaxHeight

TrueDisp of M1 vs MaxHeight estimated by Disp method, for simulated γ-rays of different energies. To decompose the Impact better, limited region of reconstructed Impact point is chosen by the selection of DispImpact between 100 m and 150 m for both telescopes.

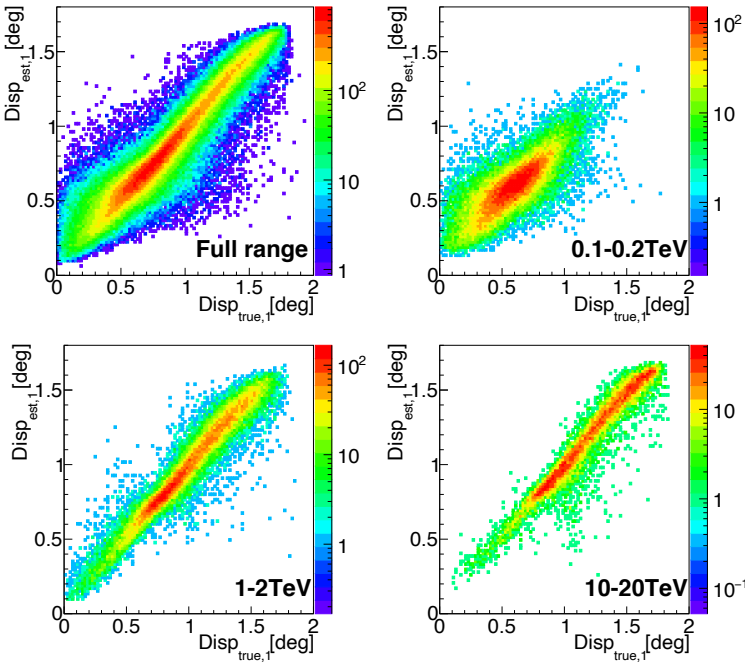


Figure 4.13: Accuracy of Disp

Estimated Disp vs true Disp of M1 for simulated γ-rays of different energies.

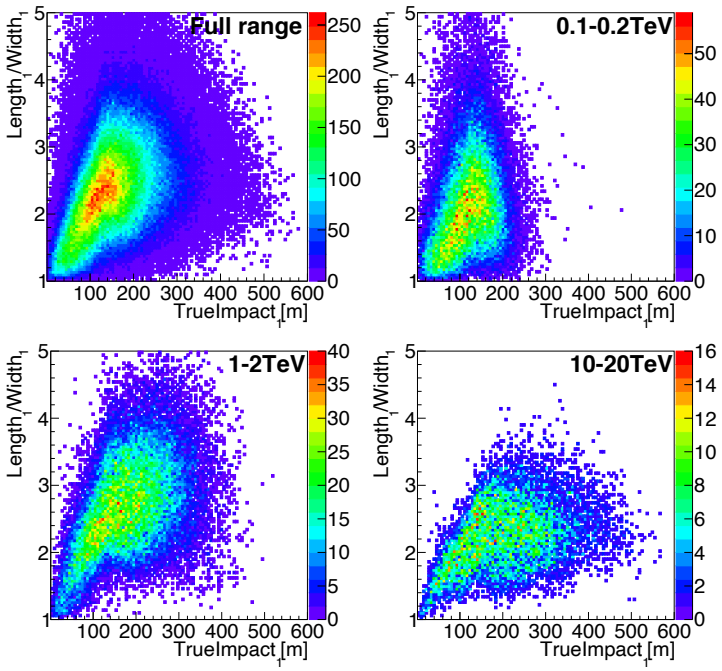


Figure 4.14: Length over Width vs Impact

Length over Width vs TrueImpact of M1, for simulated γ-rays of different energies.

Length/Width with TrueImpact in figure 4.14. They are strongly correlated, especially in small Impact range. And the correlation can be seen even in the low energy range. The small Impact is one of the ranges Disp needs a support, because the precision of Disp is affected in low value range and the range is to indicate small Impact. These variables could contribute to the energy estimation through compensating the Impact estimation.

- TimeGradient
  Among the other variables, TimeGradient is also useful parameter to estimate the parameter "Impact", as shown in figure 4.15. It is very strongly linearly correlated, especially in high energy, and in the high energy range it is the strongest variable to indicate Impact. It is also strong in high Zd observation, in which the uncertainty of DispImpact becomes too large to properly indicate Impact.

  However the correlation is smeared out in low energy, presumably because the shower image becomes smaller, and the TimeGradient cannot be calculated if the shower is so close that the light from top and bottom of the shower is no more at edges of the image.

  Moreover, it suffers from the uncertainty in which side of shower image indicates the head of the shower, consequently the confusion results in the cross shape in the relation between TimeGradient and TrueImpact, shown in figure 4.15. It is possible to determine via stereo calculation, but it still

suffers from the uncertainty to some extent. Nonetheless, the estimation of the energy is not significantly improved by the usage of the correct sign of the TimeGradient, and hence there is no need to correct for that in the strategies related to the energy reconstruction.

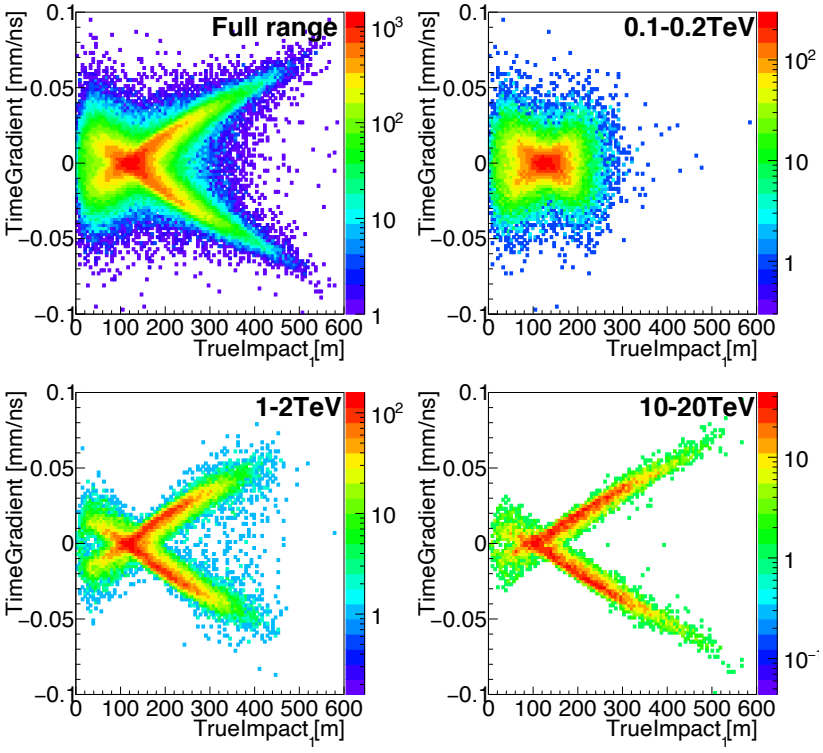


Figure 4.15: TimeGradient vs Impact

TimeGradient, namely the arrival time gradient of the photons along the major axis of the shower image (described in section 3.7.1) vs TrueImpact, of M1 for simulated γ-rays of different energies. Due to the arrival time spread at the ground caused by the two effects shown in figure 2.15, the linear relation between TimeGradient and Impact appears as described in section 2.3.4. The cross shape is due to the confusion of not knowing which side of shower image is the incoming direction. The solution to successfully determine the sign of TimeGradient is described in section 4.4.6.

The other variables related to energy estimation

Other than the variables described above, the most important input variable to include is Pointing Zenith distance angle (Pointing Zd). The characteristics of the shower images change dependent on the pointing direction, especially in the high Zenith distance range (Zd >50).

I include also the additional variables, Cherenkov density, Cherenkov radius and CosBSangle. They are migrated from LUTs-Erec. Although their contribution to the energy estimation is marginal, the redundancy would to some extent contribute to the variable randomization in the split in generating the RF, thereby to suppressing the correlation between the trees. For all these vari-

ables, Classical values are adopted since the contributions to energy estimation were not large, while Classical values are preferable to avoid systematic effects. The detail is discussed in section 4.4.2.

4.4.2 Treatment of the geometrical variables

The geometrical variables have multiple ways to be used, and here I discuss the method I adopt for the energy estimation. The variables which I list as geometrical variable in the input variables for energy are Disp, Impact, MaxHeight, CosBSangle, CherenkovRadius and CherenkovDensity. They are the variables to characterise the geometrical properties of shower events. The options of different implementations come from the two different direction reconstructions and the possibility of extracting true information about the shower geometry in simulation. The two ways of direction reconstructions are Classical method and Disp method, which results in two kinds of geometrical variables. And the variables about which we can extract true value from simulation are Disp, Impact and CosBSangle.

Among the geometrical variables, Disp and Impact are particularly important to the energy reconstruction. First I discuss them comparing the strategy to use them between previous methods and the novel method. The other variables are discussed in the second part.

Nested use of Impact and Disp, instead of Swap

The new energy estimation adopts Impact (in a form of DispImpact) and Disp for the input variables as introduced in section 4.4.1. Since the true value of these variables can be extracted in the simulated events, it has been popular strategy to make use of them in training an energy estimator. I name this strategy "swap", because the true variable from simulation is used for the training of the estimator, while the estimated variable in the measurement of the event is given as the input for applying the estimator. By the use of true value, the estimator can learn the direct relation with the target value. Therefore one may expect that it avoids the smearing of the estimator response by these geometrical variable error. This is used for Impact in the latest strategy in H.E.S.S. collaboration [40], while MAGIC used Impact and CosBSangle with this strategy in the previous standard energy estimation, namely LUTs-Erec (section 3.8.3). I also adopted it to the RF-Erec during the intermediate versions. The history of development is summarized in section 4.4.7.

However, this strategy is vulnerable to noise, thus should be treated carefully. The accuracy of estimation of these variables is limited, and systematic effects will appear when the estimated value fluctuates significantly more than the noise level of true values in train samples. In this study, the systematic effects become important in the swap strategy for high Zd range observations. The detail is separately discussed in section E.3.1 for Impact and section E.3.2 for Disp. One can clearly see the systematic effects in both variables.

Therefore I adopt a different strategy in this study, as I name it "nest" strategy. In nest strategy, true value is not given as input variable in training, instead the Disp-method-origin geometrical variables are used just like the other input variables; the estimated value is used for both training and estimation. To make use of Disp-method-origin geometrical variables as inputs in the training of

the energy estimator, the direction reconstruction needs to be already performed for the training sets. Thus the process needs to be "nested"; The RF for Disp estimation needs to be trained and applied to the train data itself, before the RF for energy is trained with the training data.

However this nesting procedure also has a possibility of systematic effect, because the Disp-method-origin geometrical variables are given to the training set in the direction reconstruction based on the Disp RF, which is generated by the same training set. Consequently the noise level of the Disp-method-origin geometrical variables in the training set is still different from the real observation data. Nonetheless, there is no visible systematic effect from nesting strategy. The investigation on the real data is also done using Crab Nebula observation in section 4.5.4.

The new procedure in the integrated MARS analysis chain is denoted in detail in section 4.4.5. The contribution of these variables are shown in section C.1.

The other geometrical parameters

Although the other geometrical parameters do not seem to directly contribute, they are still taken in the input variables to keep the trees better decorrelated to each other by maintaining good amount of input variables. To avoid the risk of the systematic effect coming from different noise level between the training sets and the data to apply, Classical method instead of Disp method is chosen.

4.4.3 Target Output and Size as logarithmic form

The target value, namely the true γ-ray energy, is fed to the training process in RF in the form of logarithm base 10. The aim is to deal with the many decades of the energy range, performing geometrical mean of the target values of the events in the last node to determine the estimation value. Correspondingly in the estimation process, the estimated value by RF is computed back to the value as power of 10.

The effect of this strategy is clearly seen in figure 4.16. In case of the estimation of energy as linear target, it suffers from the constant bias and larger energy resolution and standard deviation. And Size is also set to logarithmic scale, to keep the linear correlation in the linear scale between Size and energy as shown in figure 4.7.

4.4.4 Energy RMS

As the target is modified to be logarithm, the RMS of the prediction is also in logarithmic scale. To be available for a health check of estimation, I implemented the EnergyRMS, rescaling the prediction RMS[6] . The estimated energy $\hat{E}$ is calculated from the predictive value $\hat{y}$ as $\hat{E} = 10^{\hat{y}}$ and I define the RMS of the energy $\sigma_{\hat{E}}$ as below, calculating from the RMS of the prediction $\sigma_{\hat{y}}$.

$$\begin{aligned} \sigma_{\hat{E}_+} &:= 10^{(\hat{y}+\sigma_{\hat{y}})} - 10^{\hat{y}} \\ \sigma_{\hat{E}_-} &:= 10^{\hat{y}} - 10^{(\hat{y}-\sigma_{\hat{y}})} \\ \sigma_{\hat{E}} &:= (\sigma_{\hat{E}_+} + \sigma_{\hat{E}_-})/2 \end{aligned} \tag{4.4.3}$$

[6] See details in section 4.3.2, and a case study in section E.1

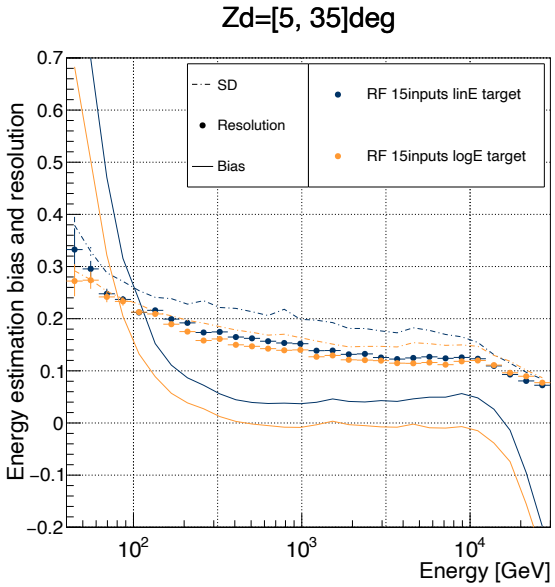


Figure 4.16: The performances of the RFs with linear target and logarithmic target

The performances of the energy reconstruction by the two RFs are plotted. "LinE target" is the RF with the target in linear form, and "logE target" is in logarithmic form. The bias (solid line), resolution (the filled circles), and the standard deviation (SD, dashed line) as a function of energy are drawn all together. The definition is explained in section 4.1.3.

4.4.5 Training process of the energy RF

Update on the analysis chain by using the nested RF

The input variables include the Disp-based geometrical variables. Therefore, as introduced in section 4.4.2, the train samples must already have the variables before being used for training the RF for energy. This requires an additional step to add Disp-based geometrical variables through direction reconstruction to the train data set. Therefore the RF for the Disp must be already constructed and applied to the train data set before the generation of the RF for energy estimation. Namely, the generations of the RFs needs to be "nested". The updated analysis flow is shown in figure 4.17.

Data cleaning on the training set

For generating the RF, the cut conditions are minimized to avoid unbalanced cleaning. Since the RF tries to average the characteristics from entire population, the data cleaning can cause biases. The removed events for training are only the events that failed direction reconstruction with Disp method.

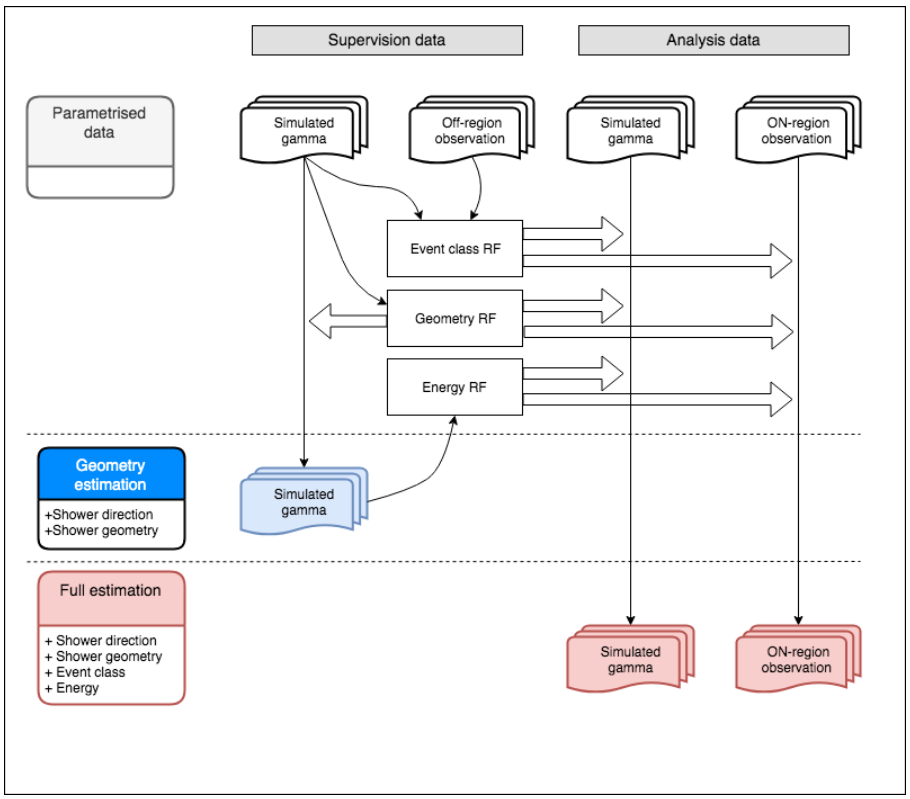


Figure 4.17: The updated analysis chain to obtain the reconstructed event properties

The two sets of data as "Analysis data" are the data that needs to be analyzed, while there are another two sets of data as "Supervision data", which are used to train RFs. A thin arrow shows flow of data and a thick arrow represents the application of a RF to a set of data. The two sets of supervision data are used to train the RFs for three event properties. In the nested RF strategy, the train samples for Energy RF need the Disp-method-origin geometrical variables, therefore the geometry RF (Disp RF) must be applied to the simulated γ-ray, before the energy RF is trained.

Forest-growing parameters

As discussed in section 4.2, RF does not require any elaborated tweak to optimize the performance. Thus the forest-growing parameters (section 4.3.2) do not affect the performance significantly, as long as the values are reasonable. In this study, the random selection of the variable is repeated 4 times in the search for the best split, taking into account that the total number of variables is 18. The number of trees of 150 and minimum node size of 3. In the implementation to the MARS software package (see section 4.4.8) they are configurable to meet the computational resource and demand for the precision. However RF performance is robust in general, and there is no significant difference. The performance study on forest-growing parameters is discussed in section D.1.

Importance of the variables (result of training)

As discussed in the section 4.4.1, the characteristics of the shower image changes significantly dependent on the Zenith distance (Zd). Correspondingly, the importances of some variables change as shown in figure 4.18. The changes of the importances of the variables can be interpreted in the following way.

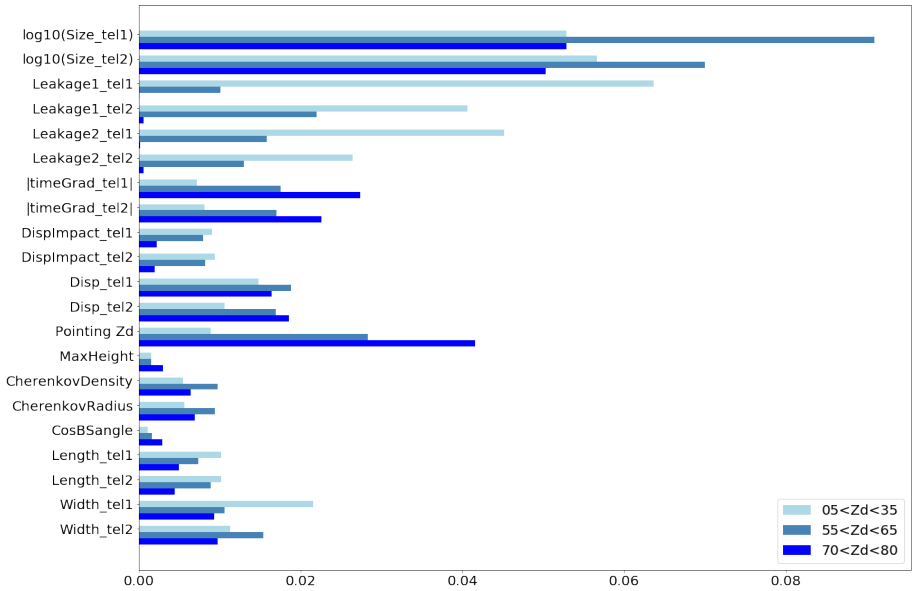


Figure 4.18: Variable importances in the energy reconstruction

The "importance", calculated as eq.(4.3.7), are shown for the individual variables in the three Zd ranges.

As the Zd increases, the shower becomes distant, and the shower image captured becomes small. The entire shower image will be contained in most of the events, therefore importance of Leakage drops. On the other hand, due to the target energy range shifting to high energy via higher energy threshold and higher collection area in high energy, the TimeGradient becomes a distinct feature.

The distant location of the shower in high Zd is also the reason why the Impact becomes less important. Since the impact point is calculated tracing from incoming direction through the shower maximum location, the longer distance to the shower maximum gives larger fluctuation to the impact point, hence the larger fluctuation to Impact. In the high Zd range, the TimeGradient becomes one of the most important variables.

And above all, the importance of Zd itself changes as the observing Zd is large. The importance is larger in the higher Zd. This reflects the small change of Zd already changes the characteristics of the shower image significantly, as the thickness of the atmosphere increases roughly $\propto 1/\cos(Zd)$.

4.4.6 Improvement of Disp

The new Disp estimation

In addition to the energy estimation, I improved the RF for Disp estimation by updating the input variables. Based on the physical properties of EAS, I improved the accuracy of Disp with two variables.

One is Leakage. Leakage was not used in the previous standard Disp RF. Leakage conveys the information of how much the image would be truncated because of the limited field of view of the camera. Thus Leakage can help with compensating the missing part of the shower image, e.g. Size and Length. Indeed, Size and Length are correlated to Disp and they are the variables which suffer the most from the truncation of image.

The other is TimeGradient. In the previous standard Disp RF, TimeGradient is used only in a form of absolute value, thus the value is degenerated, especially when the value is small. The removal of sign information is necessary to the TimeGradient, because it is calculated based only on the camera coordinates, not taking into account the shower incoming direction. I give a detailed explanation together with the solution to determine the sign of the value in the following paragraphs.

First of all, let us formulate the difference of the TimeGradient calculation between the consideration of incoming direction or not. For a given point X in camera coordinate, we represent a vector $\mathbf{X}$ from origin of the camera coordinates, denoted as $\mathbf{X} = (x, y)$. In this notation, let $\mathbf{C}$ be the CoG calculated as one of the Hillas parameters. The calculation also determines the image axis as linear function in the coordinate, namely $y = ax + b$, where a and b are constant. The gradient a can be converted to a vector parallel to image axis, $\mathbf{a} = (1, a)$. Being normalised, the unit vector $\mathbf{e}$ can be denoted as $\mathbf{e} = (\cos\theta, \sin\theta)$, where θ is the angle between horizontal axis and the image axis.

Now we convert a point on the image axis X, into the scalar to denote the position on the image axis. It is a projection of $\mathbf{X} - \mathbf{C}$ onto the 1 dimensional space with unit vector $\mathbf{e}$ along the image axis.

$$\xi = (\mathbf{X} - \mathbf{C}) \cdot \mathbf{e} \tag{4.4.4}$$

where ξ is the variable to denote the position. TimeGradient is based on the arrival time distribution of each pixel to the projected position on the image axis, namely ξ.

The problem is the choice of the projection vector. We want to know the time profile of the position along the shower image axis, as a function of the angular distance from the incoming direction. If the shower image develops in opposite side of the incoming direction with respect to the direction **e** points, the profile suffers from the flipped tendency.

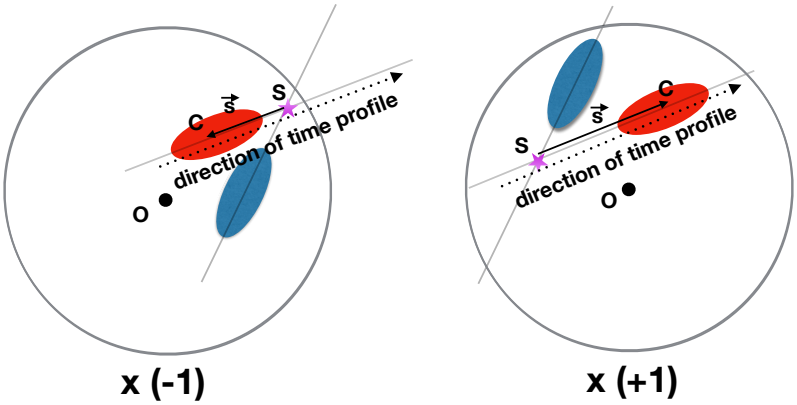


Figure 4.19: Assignment of sign to TimeGradient

Schematic views of the superposed camera images showing two cases resulting in the different time profiles. The biggest circles are the camera field of view, whose center, O, is the origin of the camera coordinates. S is the incoming direction of the γ-ray, and the two ellipses in red and blue are the shower images seen by the two telescopes. We focus on the red ellipse, whose CoG is C.

Figure 4.19 shows the two possible directions of time profile and the solution. Multiplying -1 to the TimeGradient the profile towards wrong direction, the sign of TimeGradient yields proper information. For judging the direction, the vector $\mathbf{S} - \mathbf{C}$ can be used.

$$\mathbf{e} \cdot (\mathbf{C} - \mathbf{S}) = \begin{cases} \geq 0 & \text{TimeGradient} \times = (+1) \\ < 0 & \text{TimeGradient} \times = (-1) \end{cases} \tag{4.4.5}$$

The result of the sign assignment to the simulated γ-rays is shown in figure 4.20. The left figure is after the correction with the true incoming direction for **S**. The x-shape which was present in the row distribution, shown in figure 4.15, is not seen any more. For the real data, the incoming direction is impossible to know, therefore we can only apply the classical incoming direction (the intersection of the shower axes). The right figure is the result with the correction using the incoming direction by Classical method. It still contains the events of confusion in the head-tail decision, however the majority of events are corrected with remaining mismatch events only in the order of 10 %.

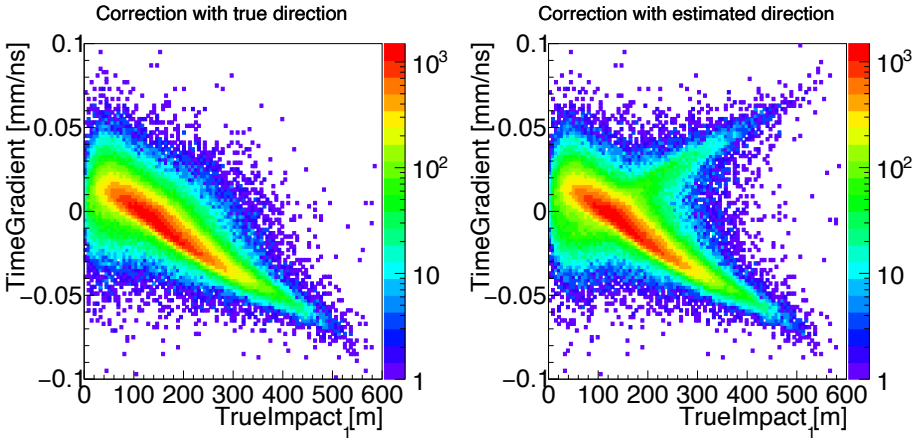


Figure 4.20: Corrected TimeGradient vs Impact

The two-dimensional histograms of simulated γ-rays as TimeGradient vs TrueImpact. The TimeGradient is corrected using the incoming direction. The left is the correction with the true incoming direction, while the right is the correction with the estimated incoming direction using Classical method (axes crossing). See also the relation before the correction shown in figure 4.15.

The evaluation of improvement

I evaluate the accuracy of Disp estimation for each event, by the sum of the errors of two Disp estimations.

$$\Delta\mathrm{Disp} = (\mathrm{Disp}_{1,\mathrm{est}} - \mathrm{Disp}_{1,\mathrm{true}}) + (\mathrm{Disp}_{2,\mathrm{est}} - \mathrm{Disp}_{2,\mathrm{true}}) \tag{4.4.6}$$

This is because the purpose of evaluating Disp is to discuss the improvement of direction reconstruction, and the Disp method for direction reconstruction is averaging of the two estimated positions. For example, if Disp for one image is longer than true value and another is shorter, it still keeps high probability of correct reconstruction to some extent.

In the statistical evaluation over the distribution of ΔDisp, there are two properties to focus; the bias which indicate the systematic deviation from the true value, and the RMS which indicate the outliers. The bias and RMS of ΔDisp are defined as

$$\mathrm{DispBias} = \frac{1}{N}\sum_{X\in\mathrm{Disp}_{\mathrm{test}}}(\Delta\mathrm{Disp}) \tag{4.4.7}$$

$$\mathrm{DispRMS} = \sqrt{\frac{1}{N}\sum_{X\in\mathrm{Disp}_{\mathrm{test}}}(\Delta\mathrm{Disp})^2} \tag{4.4.8}$$

The result is shown in figure 4.21. The improvement of Disp is seen for entire range of Disp, as the bias closer to 0 and the RMS also closer to 0. The improvement is higher at large Disp. The improvement evaluated in energy ranges shows around 0.1° improvement of bias and also around 0.1° improvement of RMS reduction in the high energy range. This results in better direction

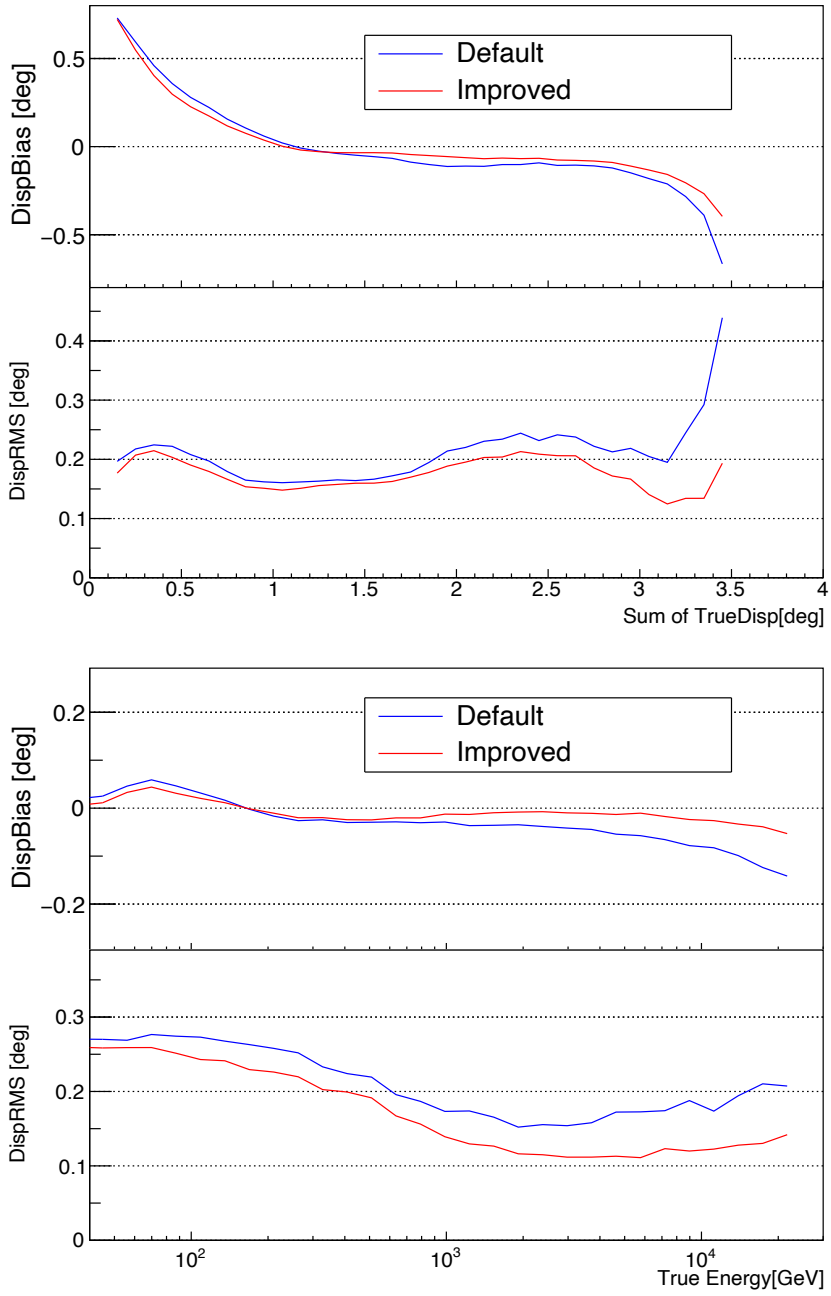


Figure 4.21: Performance in the Disp reconstruction

The two figures show the accuracy of the Disp estimation as the function of TrueDisp (top) and TrueEnergy (bottom), comparing the default Disp RF (blue) and the improved Disp RF (red). In each plot, the upper frame depicts the bias and the lower depicts the RMS, whose definitions are noted in eq.(4.4.7) and eq.(4.4.8) respectively.

reconstruction, which has an impact in the determination of signal region in the θ^2 plot.

The figure 4.22 shows the comparison of the θ^2 plots in low Zenith distance, together with the fraction of the gain of events by the improvement. The plots show that several percent of gain is expected[7].

Not only the gain in the number of the events, but also the better Disp leads naturally to a better energy estimation.

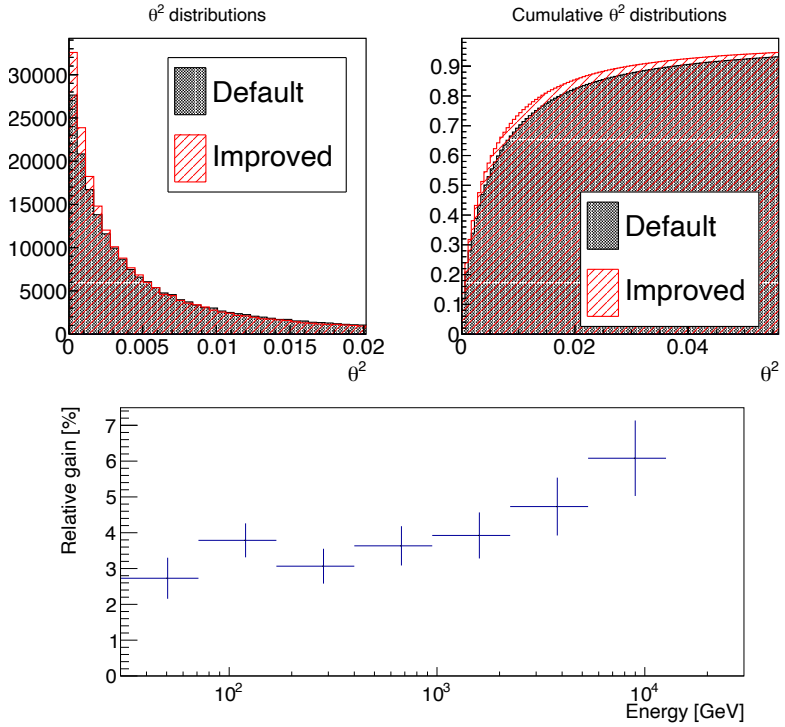


Figure 4.22: Comparison of the θ^2 distributions for the classical and improved Disp reconstruction

The classical and improved Disp estimations are compared in the reconstructed directions of standard simulation data for low Zenith distance (Zd=[5,35] deg). The top left panel is the comparison of the θ^2 distributions for all the events between the classical and improved Disp reconstructions. The top right panel is the cumulative distributions of the θ^2 distributions, normalised by the total number of events. The bottom panel is the relative gain in the different energy ranges of true energy. The calculation of relative gain is based on the number of events at the cut position. The cut position is determined by the standard procedure with efficiency of 80 %, which means that the cut position is the larger value from either at the point 80 % of the MC γ-rays are collected or 0.01 deg^2.

[7] In the standard analysis, the cut applied for θ^2 is based on the γ-ray efficiency, explained in section 3.10.3. The θ^2 cut position moves according to a given efficiency, which is the fraction to cover the simulated γ-rays after the cut. In this sense, the improvement to evaluate is the reduction in background events due to tighter θ^2 cut. However the efficiency cut sets the minimum cut value = 0.01 [deg^2] and typically the cut value reaches the minimum for the energy above 1 TeV. Therefore I evaluate the improvement by the increase in the number of events.

4.4.7 Summary of the input variables and RF-Erec versions

Compared to LUTs-Erec, the differences of input variables for the latest version of RF-Erec are

- "Nest" strategy of DispImpact instead of "Swap" strategy
- The additional use of TimeGradient, Width, Length and Disp,
- Using Classical-CosBSangle instead of swap use of Disp-CosBSangle.

Table 4.2 summarizes the list of input variables and their ways of application for the energy estimation. In the table, I also show the list of variables for LUTs-Erec, as well as all the intermediate versions of RF-Erec. They had been already released in the past as part of the MARS software package and used as official program, and were recently used also in scientific publications.

The updates underwent mainly due to the change of the geometrical variables. The difference is summarised in detail in section 4.4.2. In the latest versions, the upgrades also involve the image parameters.

Below, I summarize the history of updates together with brief justifications.

- **ver. 1 (released in May 2017)**
  The set of variables does not include any geometrical variable by Disp method, because I adopted the simplest set of variables in the picture of the shower. It already showed a better performance than LUTs-Erec.
- **ver. 2 (released in Jan 2018)**
  DispImpact was added into the input variables, which improved the performance especially in lowest energy range. I adopted swapping strategy for the DispImpact, following the strategy adopted in LUTs-Erec. However the Swap impact introduces very significant systematic effect in High Zd, which was one of the motivation to the next version. The detail of the effect is described in section E.3.1.
- **ver. 3 (released in May 2018)**
  Disp was added into the input variables, which improved the energy resolution in the entire range. The Disp did not show obvious systematic effect as Impact, thus it was introduced using swapping strategy. On the other hand, to remove the systematic effect from the swapping DispImpact, I introduced two subversions dependent on the Zd range (for more information, see section D.3).
  - **ver. 3.1 for Low Zd**
    Disp was added in swap strategy. DispImpact was kept in swap strategy because it does not show visible systematic effect.
  - **ver. 3.2 for High Zd**
    While Disp was added in swap strategy, DispImpact was removed to deal with the heavy systematic effect and instead ClassicalImpact was listed again. The parameter Length was added because it slightly improves the resolution in high Zd range. This version performs well but suffers from systematic effect when the analysis cut condition is relaxed. The detail about the systematic effect is reported in

section E.3.2, and the proper usage to avoid the effect is in section D.3.2.

- **ver. 4 (Intermediate version, not to be released)**
  The nest strategy to determine the parameters Impact and Disp is introduced to remove the systematic effect seen in the swap strategy. The strategy also reduces the outliers in the estimation error distribution in Low Zd range, as well as keeps comparable levels of the bias and resolution.

  In addition, Length and Width are added because I confirmed in the comparison between MC and Real data that the distributions of Length and Width do not show obvious differences (see the detail in section 4.5.1). I evaluated the shift of the energy estimation when one of the variables is shifted, and noted the impact on the energy estimation due to the small scale difference would be within the energy resolution. The details are described in section 4.5.2.

- **ver. 5 (Final version, to be released)**
  I improved Disp estimation as described in section 4.4.6.

Table 4.2: The variables used in the various energy estimators.

variables		RF-Erec v1	v2	v3.1	v3.2	v4,v5	LUTs-Erec
Size	(tel.1/2)	•	•	•	•	•	•
Leakage	(tel. 1/2)	•	•	•	•	•	•
Impact(tel.1/2)	Classic	•			•		
	Swap		•	•			•
	Disp					•	
Time Gradient	(tel.1/2)	•	•	•	•	•	
Width	(tel.1/2)					•	
Length	(tel.1/2)				•	•	
Disp(tel.1/2)	Swap			•	•		
	Disp					•	
CosBSangle	Classic	•	•	•	•	•	
	Swap						•
Zd of telescope pointing		•	•	•	•	•	•
Classic-Cherenkov Density		•	•	•	•	•	•
Classic-Cherenkov Radius		•	•	•	•	•	•

4.4.8 Implementation in the MARS software package

I have implemented the novel methodology into the standard MAGIC analysis software package, mentioned in section 3.3. The technical summary of the implementation is reported in Appendix B. This package is now provided as part of the standard data analysis framework, and it takes over the previous LUTs-Erec for the energy reconstruction. Thereby the MAGIC collaboration can benefit from the methodology. I also gave lectures on the usages and tips in the MAGIC software schools and MAGIC collaboration meetings.

4.5 Validation

In this section I discuss the validity of the novel energy reconstruction method (RF-Erec-v5, hereafter simply RF-Erec). First I show that RF-Erec performs consistently on simulated γ-rays and real γ-rays in section 4.5.1. They are almost consistent, apart from a small discrepancy. Then I discuss the effect of such small scale systematic effect on the reconstruction of energy in section 4.5.2. I also show event-wise comparisons, between LUTs-Erec and RF-Erec, and between the information from two telescopes, in section 4.5.3. In section 4.5.4, I show that the spectrum of the reference source, the Crab Nebula, using the novel energy reconstruction, agrees well with the past results. The agreement is good also for the spectra obtained from the observation data in different Zd ranges.

4.5.1 Variable distribution

Albeit we cannot know the true energy of the observed γ-ray, it is possible to check the consistency of the energy estimation between the simulated γ-rays and real γ-rays in observation data.

If the energy reconstruction works consistently in both the simulated γ-rays and the real γ-rays with the same energy spectrum, the measured variables of events should distribute in the same way in a given energy. This comparison is possible only if there is a bright source that allows us to reliably remove the contamination of background events. Fortunately, there is a suitable data set: the large flaring activity of the blazar Mrk421 in April 2013, which delivered the highest VHE γ-ray fluxes (during hour and day timescales) recorded with MAGIC to date. The VHE fluxes were larger than ten times that from the Crab Nebula, the bright standard candle source for γ-ray instruments.

In the following, the Monte Carlo (MC) simulated data of γ-rays is compared to the (Real) observation data.

The basic idea

The variables distribute differently over energy, and this is how it is possible to judge the energy of the γ-rays from the set of variables. In other words, for a γ-ray with a given energy coming from a given direction towards a given position on the ground, the measured value of the variables should follow the correspondent probability distribution. Therefore the accumulated distribution of measured values for γ-rays with given energy must be the same between real and simulated γ-rays, if they come from the same direction and arrive uniformly on the ground. This comparison should work in the estimated energy in the same way.

There are still some issues to carefully take into account as follows.

- Astrophysical sources emit following roughly power-law spectra with negative spectral indices, not to mention that there is no mono-energy source. Therefore the samples with the estimated energy in a certain range are the mixture of different true energies, often dominated by lower true energy events because of steep power-law nature. To take into account this effect, the comparison between the MC data and the real data requires

the adjustment of the spectrum, so that the mixture of the true energy in the estimated energy domain becomes the same.

- The source in the sky moves, and the relation between the true energy and observed variables changes correspondingly. Therefore the Zenith distance distribution must also be taken into account.
- The real γ-ray event is embedded among the dominating background events. Therefore the distribution of entire data will end up showing the distribution of background events. However, like the usual process to extract the statistics of the γ-ray, the distribution of the γ-ray events should emerge by performing the on-off subtraction.
- The flux of real γ-rays we observe is typically very low and dominated by the overwhelming flux of the cosmic-ray background events. Therefore the observation data must contain very high flux of γ-ray.

Therefore we compare the distributions of observed variables for γ-ray events dependent on the energy between simulated events and the real observation data, namely the two-dimensional histogram of the estimated energy and the variable. In the following, we discuss how to proceed with the issues above.

The (Real) observation data

The data set selected for the these studies is the observation of Mrk421 taken in April in 2013. Figure 4.23 shows the summary of the data used in this study.

The source exhibited the highest flux level, persistent over several days, among all the TeV γ-ray sources. I select the first four days in the high flux term, April 12, 13, 14 and 15 in 2013. The total observation time is $\sim$ 20 hours. For these four days, the integrated flux $>$ 200 GeV was the highest, varying from $1 \cdot 10^{-9}$ to $2 \cdot 10^{-9}$cm^{-2}s^{-1} , which corresponds to around 5 to 10 C.U.. The data used here was taken at low Zd, and the weather condition was very good.

After the event reconstruction, the cut conditions applied are Size > 50, θ^2 efficiency $= 0.9$ and $Hadronness$ efficiency $= 0.9$. The spectrum of γ-rays for the entire data is estimated by forward folding. The spectral model applied is power law with semi-exponential cutoff which gave the best fit result among the usual assumptions with the reasonable probability, $\chi^2/n.d.f. = 13.2/16$ (p-value $= 0.66$).

$$f(E) \propto \cdot \left(\frac{E}{E_0}\right)^{\alpha} \cdot e^{-(E/E_c)^{\beta}} \tag{4.5.1}$$

where E is the energy in GeV, $E_0 = 387$ GeV, $E_c = 1044$ GeV, α= -1.8 and β=0.65.

To construct the variable distributions of γ-ray events, the main concern in the real observation data is the background events. The histogram needs the subtraction of the contribution from background events. Since the data is taken by wobble mode, it is possible by the ON - OFF subtraction. The events around the source direction are filled to on-histogram, and the ones around the axi-symmetric direction to an off-histogram. The residual of these two histograms

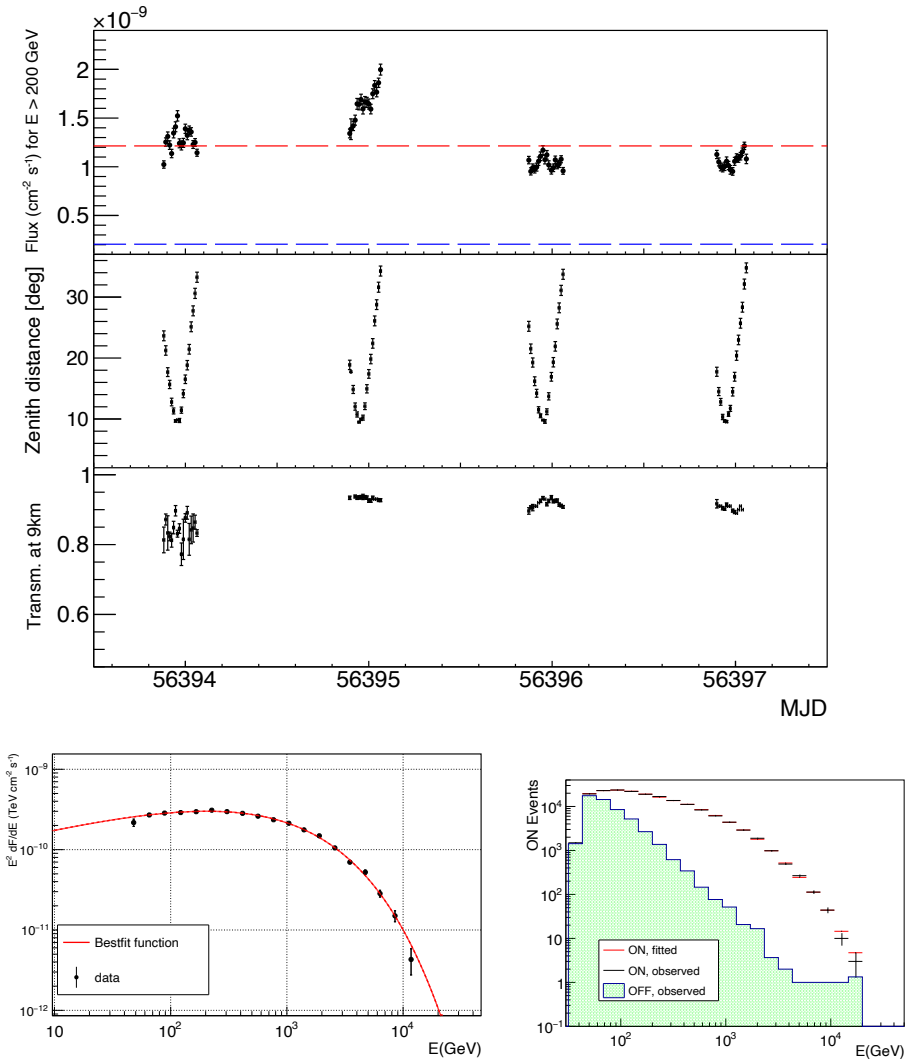


Figure 4.23: Characteristics of the Mrk421 data taken in Apr 2013

Top: The estimated source flux and the observation condition as a function of time. From top to bottom, the light curve, namely the flux vs time, the Zenith distance angle, and the transmission of the atmosphere. The red line is the average flux for all the data, and the blue blue line is the flux level of the Crab Nebula.
Bottom left: The spectrum of γ-rays estimated by forward folding. The red curve is the model function with the fitted parameters, and the points are the reconstructed flux as a function of energy.
Bottom right: The observed number of events vs estimated energy. The number of events in ON and OFF regions are plotted, together with the number of events in ON region, obtained from forward folding.

should indicate the distribution of real γ-ray events, because the distribution of background events should be the same for the axi-symmetric directions.

The cut conditions applied on both on- and off- histogram are $\theta^2 < 0.1$ and $hadronness < 0.95$. It is to avoid the unnecessary events. From the θ^2 distribution shown in figure 4.24, the criteria sufficiently covers the distribution range of signal events.

The (MC) simulated data

For the MC data, there is no need for subtraction of background, but instead the energy spectrum and the Zenith distance (Zd) distribution must be the same as the real data. Therefore the distribution of MC data is adjusted giving weights to individual events.

The weight for adjusting the spectrum is calculated from the true energy, so that the resultant distribution of the true energies for given estimated energy is equalized. The spectrum of the real data is assumed to be the spectral function obtained from the forward folding, mentioned above. On the other hand, the spectrum of the MC data can be regarded as the probability distribution on energy when it was generated, which is proportional to $E_{true}^{-1.6}$.

The weight for the Zd is calculated from the difference of the Zd distributions between observation data and the MC data shown in figure 4.24. The distributions are binned in the cosine of Zenith distance with 10 bins. For the observation data, the amount to be counted in each Zd bin is the observation time. On the other hand, that for the MC data should be characterised in the number of simulated events. Based on them, the weight for an event is assigned from the scale factor in the Zd bin where the event belongs.

To impose the same condition between the real data and the MC data, the same cut conditions as the observation data are applied. And the two weights are adjusted further so that the total number of events between the MC data and the real data (ON - OFF) become the same after the cut.

In the comparison of θ^2 distributions shown in figure 4.24, the MC data and the real data show discrepancy. The MC data has higher fraction accumulated to the smallest θ^2 bin, and the real data has higher fractions than MC data in the next bins. It indicates that the event reconstruction done for MC is better than real data. This small difference can naturally come from the difference in Point Spread Function (PSF) of the Cherenkov light collection, which is mostly from the adjustment of the reflector system.

The result

Here I show the variable distributions for the MC and the real data and compare them. I select the variables with highest contributions to estimate the energy. Leakage is omitted because the distribution is dominated by 0 and it does not make sense for the comparison. For simplicity, the distributions of the telescope wise variables are shown only for telescope 1 (M1), because the trends are similar for both telescopes.

The resultant variable distributions are shown in the top left panels for the MC data, which denote simulated γ-ray, and in the top right panels for the real data as ON - OFF of the distributions, which denote the real γ-rays, in the figures from figure 4.25 to figure 4.30. The distributions are drawn in two

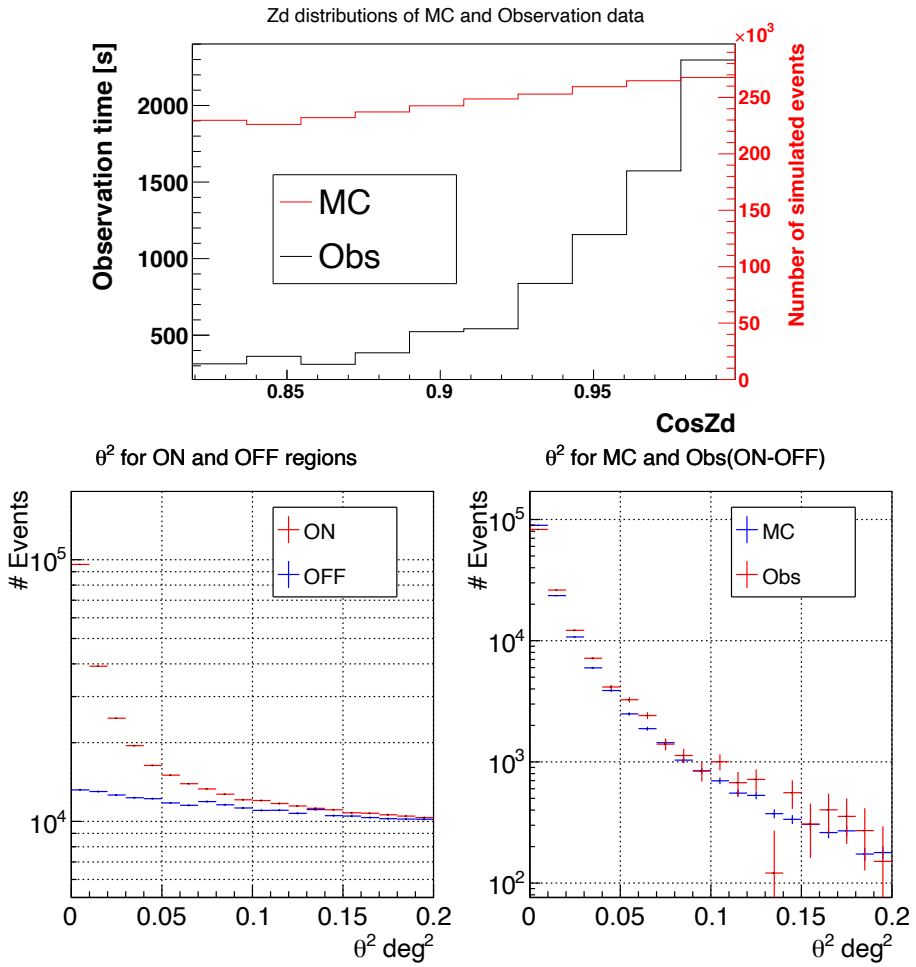


Figure 4.24: The comparison of the simulated data with the real data, in θ^2 distributions and Zd distributions

The Monte Carlo (MC) simulated data of γ-rays is compared to the observation (Obs) data to see if it reproduces reasonably the real γ-rays in the observation data. In addition to the spectral weight, the weight to take into account the Zd distribution is applied.
The top panel is the distribution of Zd for the simulated data and the observation data.
In the bottom, the performance of direction reconstruction is compared. The bottom left panel is the θ^2 distributions in the ON and OFF regions of the Obs data. The subtraction of them yields the γ-rays, that are shown in the bottom right panel. The θ^2 distribution of simulated γ-rays is shown together in the panel. The weights for the spectrum and Zd are applied.

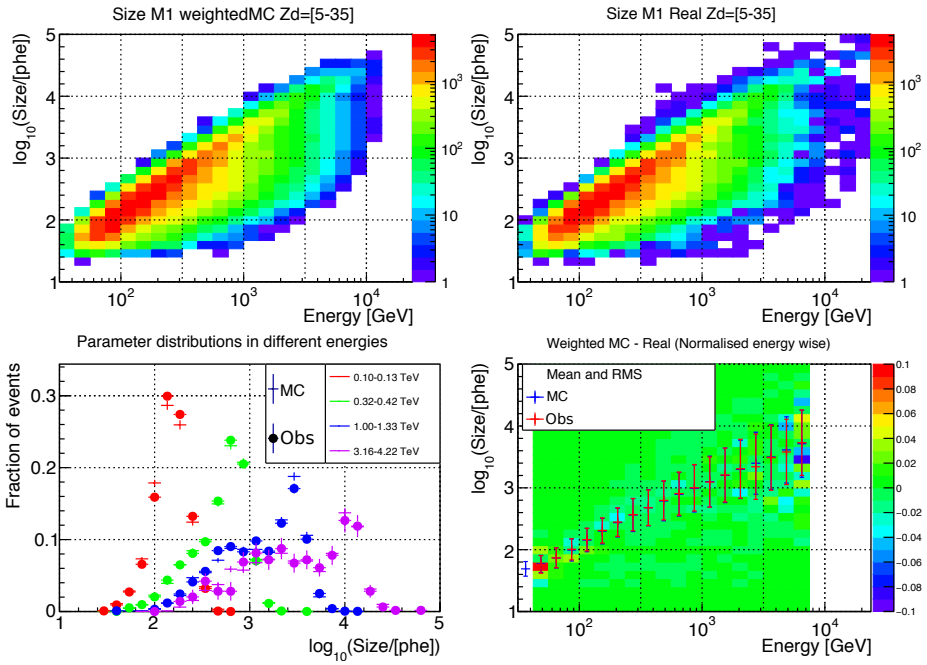


Figure 4.25: The variable distributions of Size of M1

The distributions of Size of the telescope 1 (M1) vs estimated energy are drawn in the top two panels. The top left panel is for the simulated γ-rays (MC) weighted in the true energy and Zd. The top right panel is for the observation (Obs) data as excess events of ON - OFF. The two distributions are compared in the bottom two panels. The bottom left panel is the profile histograms, for four selected energy bands. The bottom right panel is the mean and the RMS around the mean in the binned energy ranges, on top of the residual of the two histograms normalized by the total number of events in each estimated energy bin of the observation data. The energy range with less than 100 events of the excess events of the observation data is omitted.

dimensions so that the energy dependence of the variable distributions can be seen. At a glance, the distributions are very similar for all the variables between the MC and the real data.

The detailed comparisons of the two distributions are seen in the bottom panels. The bottom left panel shows the profile histograms, which are the variable distributions in four selected energy ranges; 0.10 - 0.13 TeV, 0.32 - 0.42 TeV, 1.0 - 1.3 TeV and 3.2 - 4.2 TeV. The shapes of the profile histograms are irregular, however the data points of the MC and the real data mostly agree with each other.

The bottom right panel shows the subtraction of the two histograms. The distribution of real data is subtracted from that of MC data, and is shown as the ratio to the total number of events in the real data integrated over each energy bin. All the grids with good statistics are within several percent of the total events in the estimated energy bin. The graph superposed indicates the mean and the RMS of the distribution in each energy.

All the variables show consistency between the MC data and the real data.

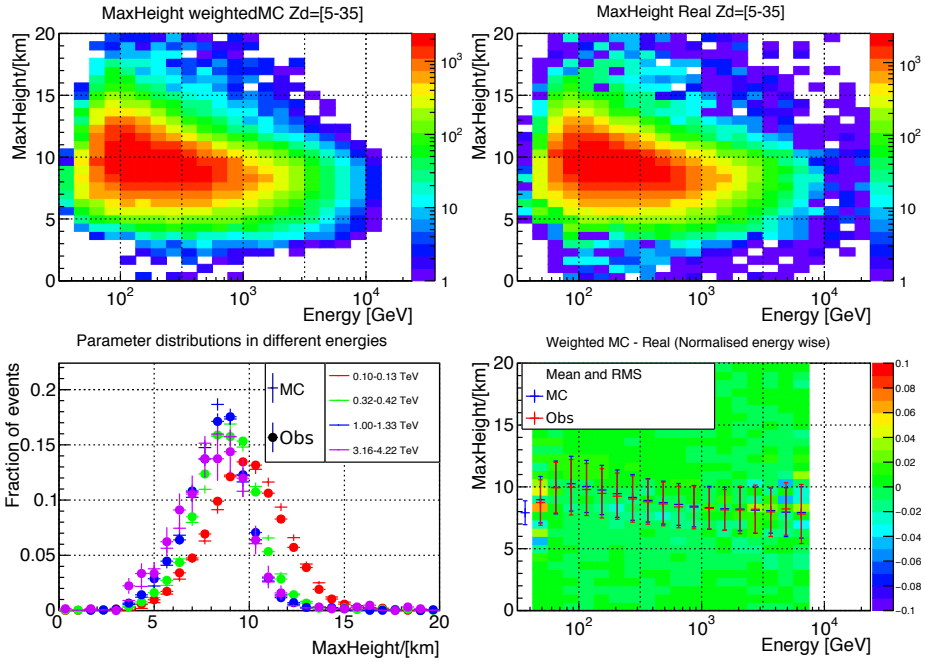


Figure 4.26: The variable distributions of MaxHeight

The explanation follows in the same way as figure 4.25.

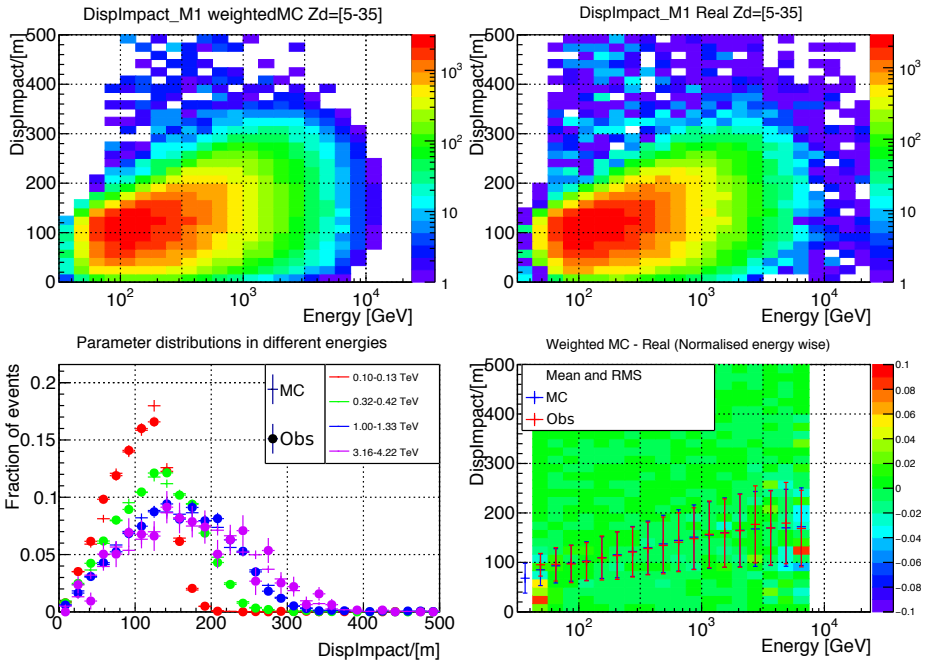


Figure 4.27: The variable distributions of DispImpact of M1

The explanation follows in the same way as figure 4.25.

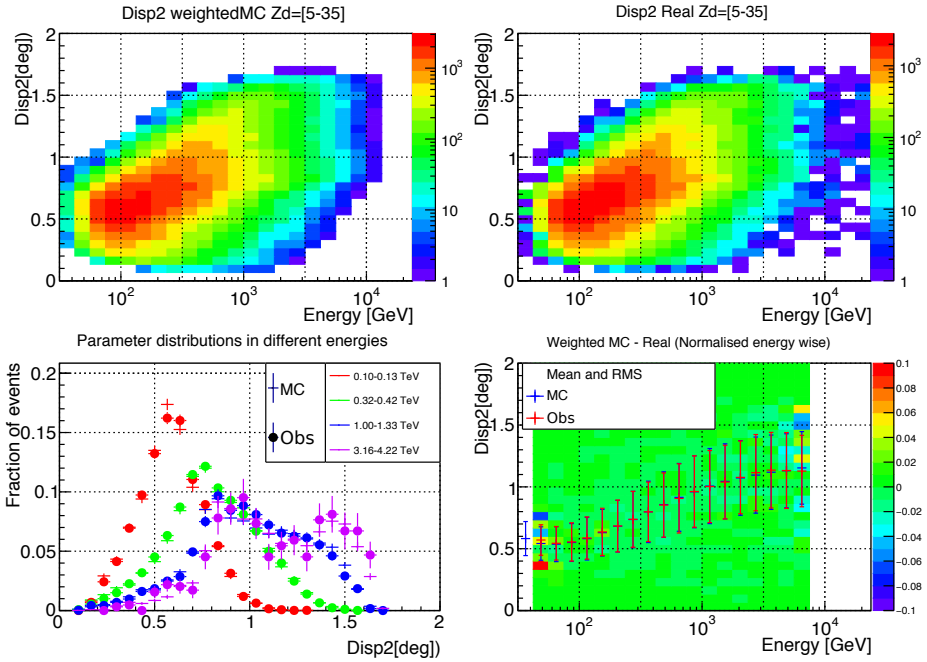


Figure 4.28: The variable distributions of Disp of M1

The explanation follows in the same way as figure 4.25.

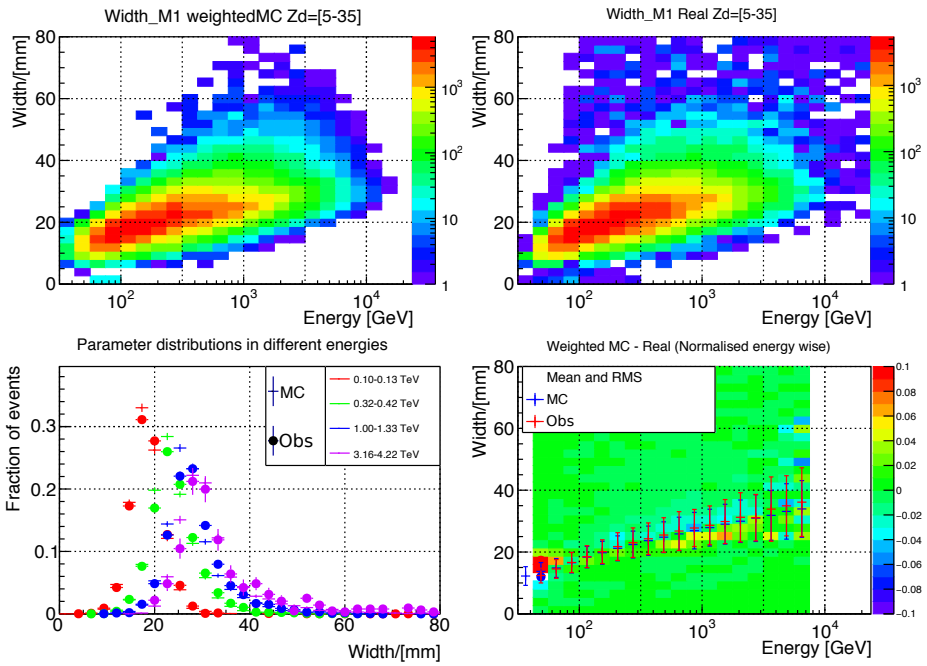


Figure 4.29: The variable distributions of Width of M1

The explanation follows in the same way as figure 4.25.

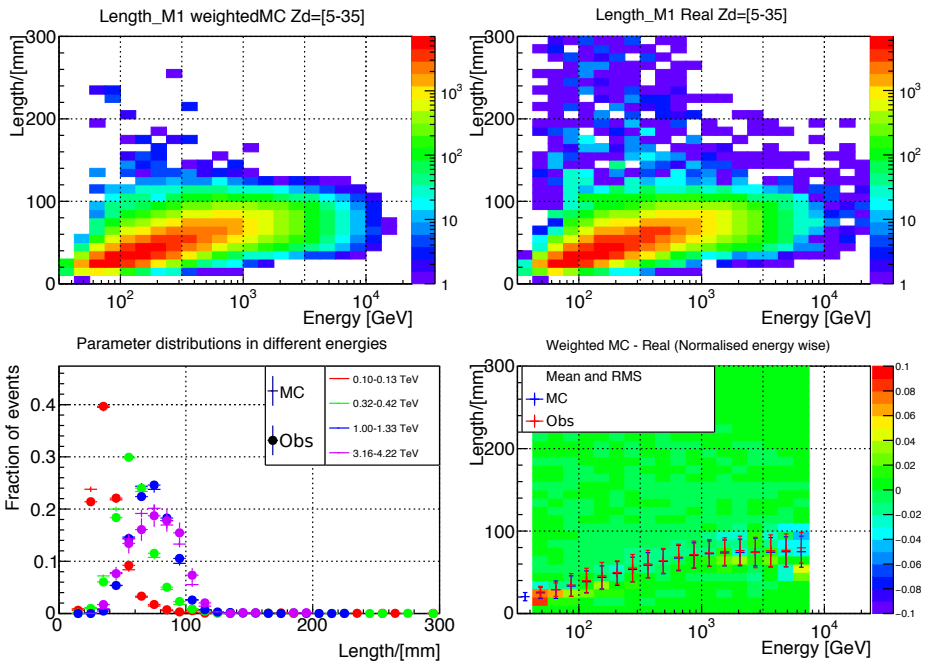


Figure 4.30: The variable distributions of Length of M1

The explanation follows in the same way as figure 4.25.

It is remarkable that the profile histograms mostly agree with each other in different energy ranges, although the shapes of the distributions are complicated and vary in energy. However some discrepancies can be seen. The discrepancies in the lowest energy and the highest energy are not a problem because they suffer from the limitation in statistics. The systematic difference seen especially in Width distribution should be discussed. The residual histogram shows it is a systematic shift; the mean of the distributions is shifted by around 1 mm.

One challenge in the MC-real data comparison is the uncertainty in the spectrum variability. To give the weight for adjusting spectrum of the simulated data, the primary spectrum of the observation data is fitted to the entire data. However the light curve shows very strong variability with the changing of flux by factor of more than two even within a night. Therefore there is no guarantee of the spectrum shape to be stable, and if the spectrum varies with time, the weight dependent on the time spent for the Zenith distance cannot accurately work. It is also supporting that the variable distribution really changes dependent on the Zenith distance weight.

Another challenge is the difference of the pointing direction distributions in Azimuthal (Az) direction. The observation tracked the source, thus the Az distribution is narrow; but the simulation data is generated uniformly in Az direction. Usually Az dependence has been neglected in low Zd observation in standard analysis, however the systematic effect might be non-negligible now, because of better precision. To confirm this possibility, dedicated simulation is needed, which is out of the scope of the work presented here.

However the most likely reason that caused the discrepancy is the difference in the PSF of the reflector system. As shown in the figure 4.24, the θ^2 distributions are different between the MC and the real data. The PSF is measured regularly, however it is not possible to have a nightly dedicated measurement. On the other hand, the PSF always changes slightly, mainly due to the deformation of the telescope structure. Therefore it is unavoidable to contain a small deviation from reality. In this study, the θ^2 distribution is better in MC data, which indicates that the direction reconstruction is performed better in the MC data. It is likely that the MC data was produced with the better PSF in the simulation. And this is consistent with the variable distributions of Width, where the Width of the MC data distributes in smaller values than that of the real data.

Above all, the variable distribution shows good agreement between simulated γ-rays and the real observation data.

4.5.2 Energy uncertainty caused by MC-Real data differences

The variable distributions reported in section 4.5.1 show good agreement between the simulated γ-rays and the real γ-rays, although a small difference can still be seen. As mentioned, the reason of the difference could be attributed to the uncertainty of the time-varying true spectrum or the different Azimuthal angle distribution between the simulation data and observation data. However an existence of bias in the simulation cannot be excluded for now, e.g., owing to the difficulty of simulating perfectly the optical PSF. This section shows the impact in the reconstructed energy if one assumes that the systematic difference of a measured variable is genuine.

The procedure

In this study I assume a simple case of systematic difference, where a variable is systematically larger by certain fraction or interval for all the events. I compare the reconstructed energies of the γ-rays between the original input variables and those with a variable shifted systematically. The comparisons are performed for different variables. I select them based on the importance to determine the energy as discussed in the variable selection in section 4.4.1. Table 4.3 shows the list of variables and the shifts used in this study. From the comparison between the simulated γ-rays and the real γ-rays the systematic difference of a variable would not be larger than those.

I select the same low Zd (between 5 deg and 35 deg) as the variable distribution test.

The result

The result of the investigation for the energy between 0.5 TeV and 1.6 TeV is reported in figure 4.31. I set this energy range because it is the most important range to most of the VHE γ-ray sources, as well as because it is the range with the best reconstruction performance. Moreover the biggest discrepancy seen between the simulated γ-ray and the real γ-ray is in Width around this energy range, as shown in the left bottom panel of figure 4.29. Also this energy range

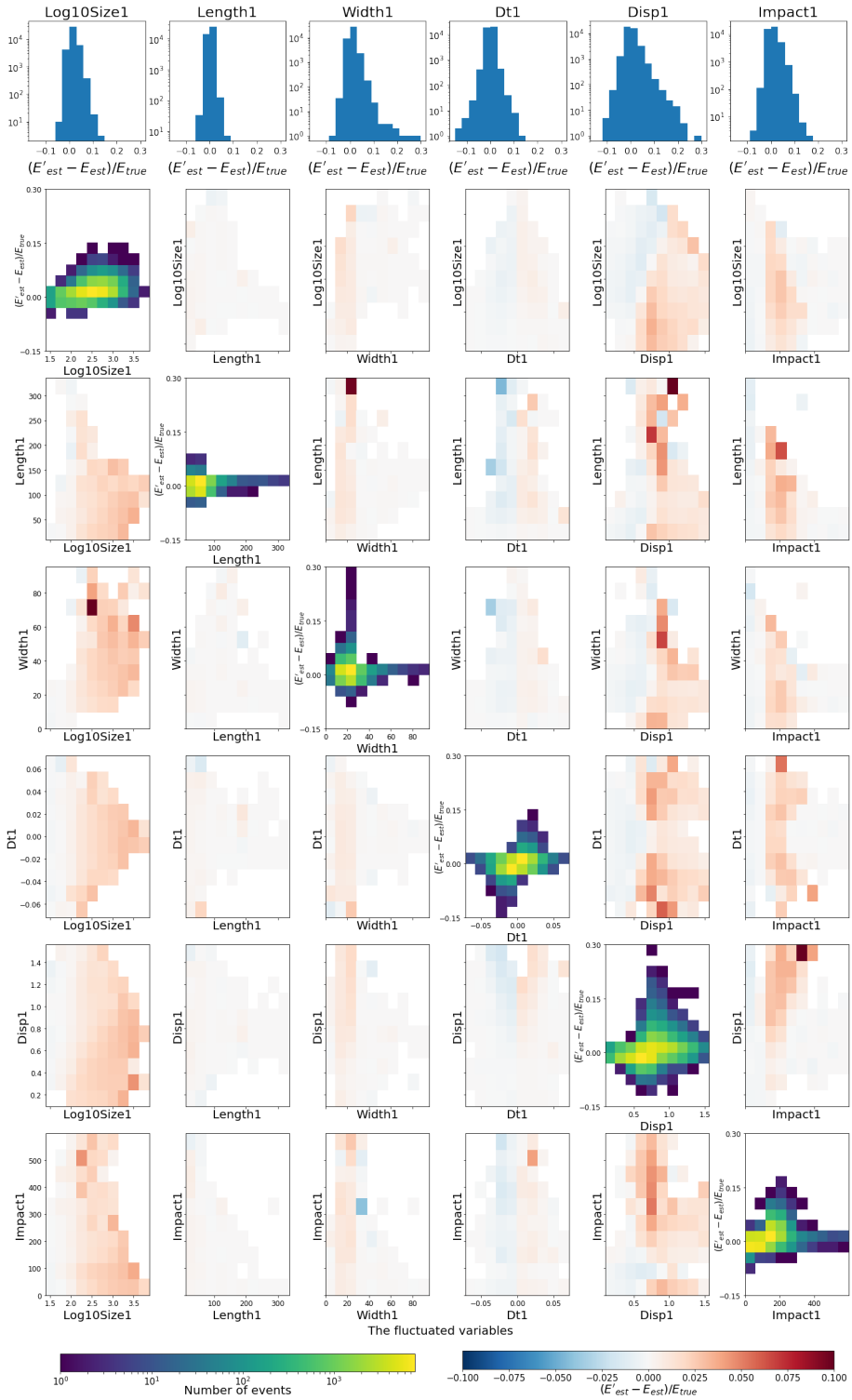


Figure 4.31: Energy uncertainty caused by MC-Real data differences

The distributions of the estimated energy (Eest) shift due to the shift of a variable. The shifted variables are all in M1, which are from left to right, Size, Length, Width, TimeGradient (Dt), Disp and Impact. The amounts of shift given are listed in table 4.3. Top row is the entire distribution, and the other rows show the pair plots, in which the distributions are decomposed to another variable. The diagonal plots are two dimensional histograms of energy shift vs. the shifted variable. The off-diagonal plots are the heat maps of average energy shift for the grid of the shifted variable and another variable.

Table 4.3: Selected variables and their shifts

Variable	Shift
Size1	+ 5%
Length1	+ 2 mm
Width1	+ 1 mm
TimeGradient1	+ 0.01 ns/mm
Disp1	+ 0.05deg
DispImpact1	+ 10 m

seems to be the most sensitive to the shift of the variables, because of the good statistics of number of photons per event and the good statistics of events.

Each column represents the shifted variable, from left to right, $Size_{\text{tel.1}}$ (shown in logarithmic form), $Length_{\text{tel.1}}$, $Width_{\text{tel.1}}$, $TimeGrad_{\text{tel.1}}$, $Disp_{\text{tel.1}}$, and $Impact_{\text{tel.1}}$. The top row reports the histograms of the shift of reconstructed energy, normalized to the true energy. For all variables shifted, most of the impact in the reconstructed energy is typically less than 10 %.

The other panels show the dependence of the energy shift on the variables in the diagonal plots, and the pair plots of heat maps in the rest, where correlations between two variables are searched. The heatmaps depict the average of the reconstructed energy shift, red for positive shift and blue for negative shift. A map is binned in two dimensions, of which the horizontal axis is the shifted variable and the vertical axis is another variable. Looking closely, the influence of the variable shift on reconstructed energy has trends.

4.5.3 Event-wise consistency checks

Comparison with LUTs-Erec

Since the LUTs-Erec has been used as the official strategy in the MAGIC collaboration during the last decade, if there is an anomaly in the distribution of the event-wise difference of estimated energy between LUTs-Erec and RF-Erec, a reasonable explanation should be given. Figure 4.32 shows the distribution of event-wise difference of estimated energies between LUTs-Erec and RF-Erec, $(\mathrm{E}_{est}^{RF} - \mathrm{E}_{est}^{LUT})/\mathrm{E}_{true}$, to the true energy in the form of logarithm of 10. The majority of the events are distributed symmetrically around 0, therefore there is no bias between the LUTs-Erec and RF-Erec. However, the outliers are not symmetrically distributed, showing the long tail only in the negative side. This seems to come mainly from the systematic effects by swapping Impact, which are discussed in section E.3.1.

Comparison between two telescopes

The information from the two telescopes should contribute equally to the energy estimation. To check this, two RF energy estimators are constructed, in which the image-based parameters of either one telescope are removed from the input variables. Compared to the estimator with the full input variables, the two estimators bias should lose the information symmetrically, thus the difference of the estimations by them should be distributed symmetrically. The result of this test is shown in figure 4.33, confirming that the distribution is symmetric.

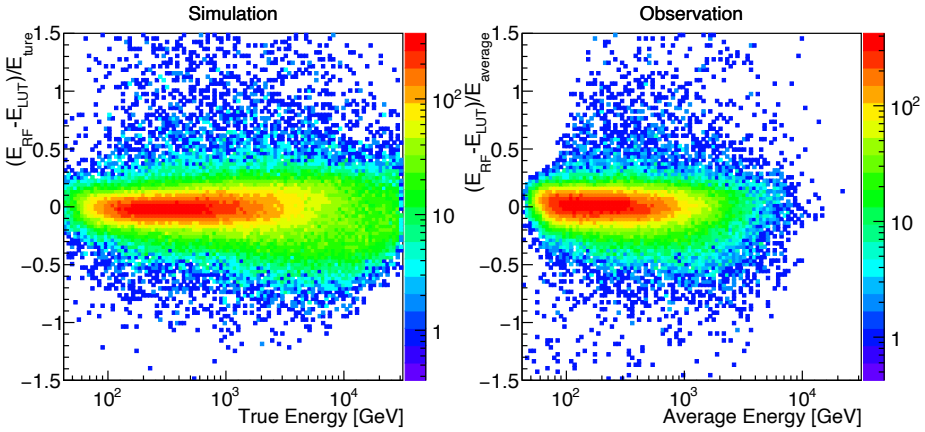


Figure 4.32: The event-wise comparison of LUTs-Erec and RF-Erec

Difference of the reconstructed energies by LUTs-Erec and RF-Erec, as a function of energy. The left panel is for the simulated γ-ray (MC data). In the vertical axis, the difference of the reconstructed energies is normalized by the true energy, and the horizontal axis is true energy. The right panel is for the observation data. The difference of the reconstructed energies is normalized by average energy, which is the geometrical mean of the two reconstructed energies. Horizontal axis is the average energy. The events are selected under the cut condition of
$0 \leq \theta^2 \leq 0.02$ & $Hadronness \leq 0.5$ & $50 \leq Size_{\mathrm{tel.1}}$ & $50 \leq Size_{\mathrm{tel.2}}$.

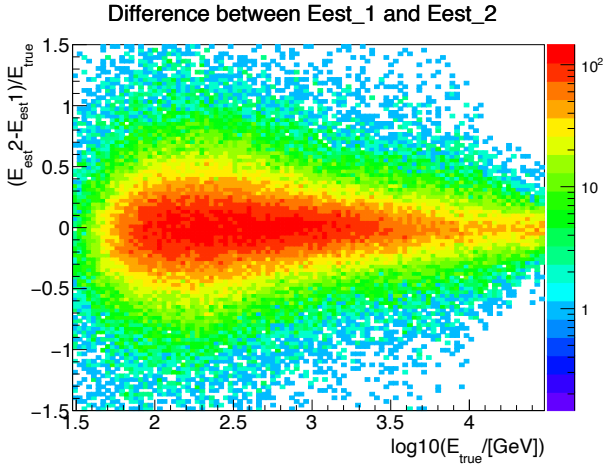


Figure 4.33: The event-wise comparison of the reconstructed energies by the RFs with limitations on image-based parameters of M1 and M2

Difference of the reconstructed energies by the RFs with limitations on image-based parameters. The two RFs are trained with reduced variables, removing the image-based parameters of either one telescope. The reconstructed energy $E_{est}1$ denotes the reconstructed energy by the RF keeping the image-based parameters of telescope 1 (M1). The difference is normalized by the true energy E_{true}. The events are selected under the cut condition of
$0 \leq \theta^2 \leq 0.02$ & $Hadronness \leq 0.5$ & $50 \leq Size_{\mathrm{tel.1}}$ & $50 \leq Size_{\mathrm{tel.2}}$.

4.5.4 Crab Spectrum

As described in section 3.12.1, the most pragmatic test to validate RF-Erec consists in determining the spectrum for our bright standard candle, the Crab Nebula, and compare it to that produced with the LUTs-Erec.

I performed this test in four Zenith distance ranges; Zd=[5,35]deg, Zd=[35,50]deg, Zd=[55,65]deg, and Zd=[70,80]deg. In this subsection I call them specifically as LZd, MZd, HZd and VHZd. The observation data used here was taken during astronomical night, with no moon in the sky or near the horizon, and in good weather, to avoid any other systematic effects. The total observation time collected are 3.6 hours, 2.1 hours, 3.7 hours and 20.5 hours for LZd, MZd, HZd and VHZd [8].

The flux points are calculated by the number of excess events divided by collection area, which was corrected for the energy migration using the migration matrix[9], collection area for true energy and the assumed spectrum obtained from the past study [16]. The cut condition was optimized. Since the source is bright, the θ^2 efficiency was increased to 85%. The hadronness efficiency was left to be the standard value of 90% for LZd and MZd, however HZd and VHZd needed it to be relaxed to be 99% because there appears the systematic effect in the hadronness distribution, due to the differences between real and MC data. The Size cut remained the same, 50 phe.

The result is shown in figure 4.34. For comparison, the spectra via the LUTs-Erec are also shown. The vertical lines are the energy threshold to properly estimate the spectral data points. For more details, see section D.4. The red dashed lines are the reference spectrum obtained from the past study.

When the spectral shape is fitted to the spectral data points, they all agree with the reference spectrum. Therefore the energy reconstruction worked well. As for the individual data points, they should be consistent with the reference spectral lines, given the assumed spectrum for the flux calculation is valid. All the data points are well aligned with the reference spectral lines, except for some deviations, which are mostly within around 2 sigma. There is an exception at around 6 TeV in the Spectral Energy Distribution (SED) by LUTs-Erec in HZd, which deviates by around 3 sigma. This is very likely due to the systematic effect induced from the swapping, because the deviation coincides with the isolated probability bin in the migration matrix, shown in figure 4.35. The mechanism is discussed in section E.3.1, and a clearer example for LUTs-Erec is seen in section 4.6.

[8] All the selected data are in the same period described in section 3.4.3. The period selected is from 2016-04-29 to 2017-08-02. This period is compatible with the telescope performance in 2019.

[9] The migration matrices used are presented in figure 4.35.

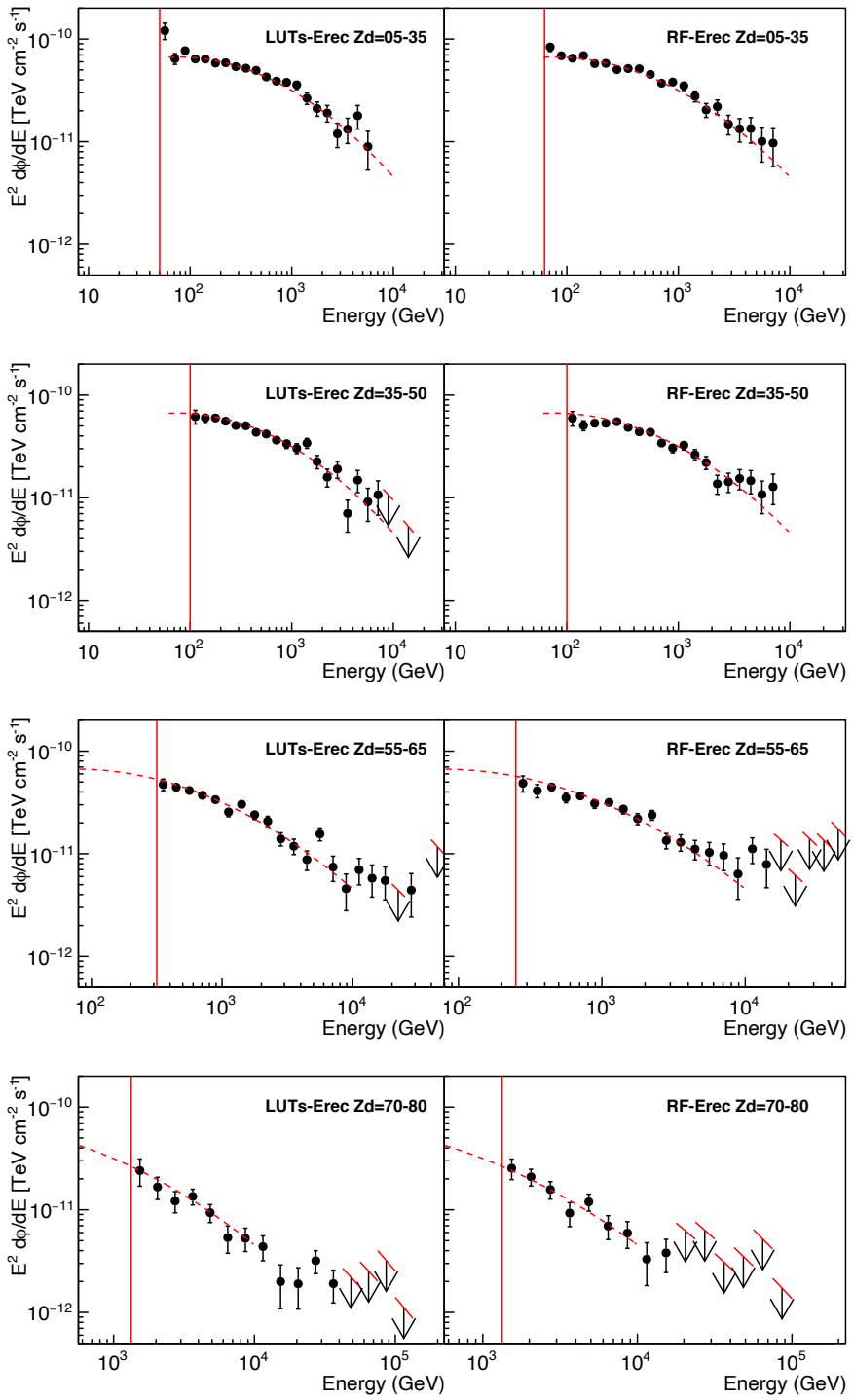


Figure 4.34: Validation of the energy reconstruction via the Crab Nebula spectra

The Crab Nebula spectra obtained with LUTs-Erec (left) and RF-Erec (right). From top to bottom, the Zenith distance range of the observations are indicated in the panels. The red dashed line is the Crab Nebula spectrum presented in past study by MAGIC collaboration [16].

4.6 Performance evaluation

In this section I discuss the performance of the new energy estimator, RF-Erec, compared with the previous strategy, LUTs-Erec. The simplest and most general way of measuring the performance is through properties related to generalization error which is introduced in section 4.1.3. I show the migration matrix as raw distribution of the estimation error in section 4.6.1, and quantify the properties in section 4.6.2.

However, these are not sufficient to prospect the effect on the result of analysis, because the goal of the analysis is to extract physical parameters from a spectrum after the accumulation of statistics. The reconstructed spectrum is deformed because the confusion of the estimated energy in individual events changes the actual distribution. And the deformation is strongly dependent not only on the reconstruction performance, but also on the spectral shape itself. Moreover, scientific interests depend on different parts of the spectrum.

For example, the slope of a spectrum changes the strength of deformation. The interest in existence or not of cutoff requires reliability to measure the highest energies. The interest in anomaly in the spectral shape like a bump or dip requires stable performance over energy.

In this study, I evaluate the estimator in two ways; the energy-wise bin reliability of the events, and the actual effect on modeling a spectrum. The energy-wise bin reliability is especially important for the high energy range, where the data runs short of statistics. For this purpose, I discuss bin purity and spillover in section 4.6.3. For the actual effect on modeling a spectrum, I evaluate the reliability on determining the spectral parameters. I discuss power-law unfolding in section 4.6.4.

Throughout this section, I use the same set data that were used in section 4.5.4.

4.6.1 Migration matrix

The migration matrix, introduced in section 3.11, is the best measure to directly asses the performance of an energy estimator. The migration matrices from RF-Erec and LUTs-Erec are shown in figure 4.35, separately in different Zenith distance (Zd) ranges; Zd=[5,35]deg, Zd=[35,50]deg, Zd=[55,65]deg, and Zd=[70,80]deg.

The migration matrices obtained with LUTs-Erec show wide spread of migration probabilities while the ones from RF-Erec have tight band of bins along the diagonal line. The tight spread indicates good energy resolution, which means that the spread of the reconstructed energies for a given true energy is small. The quantification of the properties to characterise this is discussed in section 4.6.2.

In the example shown here, the LUTs-Erec migration matrix for Zd=[55,65] deg has a cell, which tells the events of the true energy < 100 GeV have a probability above 20 % to migrate to estimated energy of about 5 TeV. These features bring difficulty in precisely unfolding an energy spectrum. This issue can cause a visible systematic effect in an unfolded spectrum, especially when the spectrum is steep. Some examples of the actual influences on the analysis can be seen in section 4.6.3 and section 4.6.4.

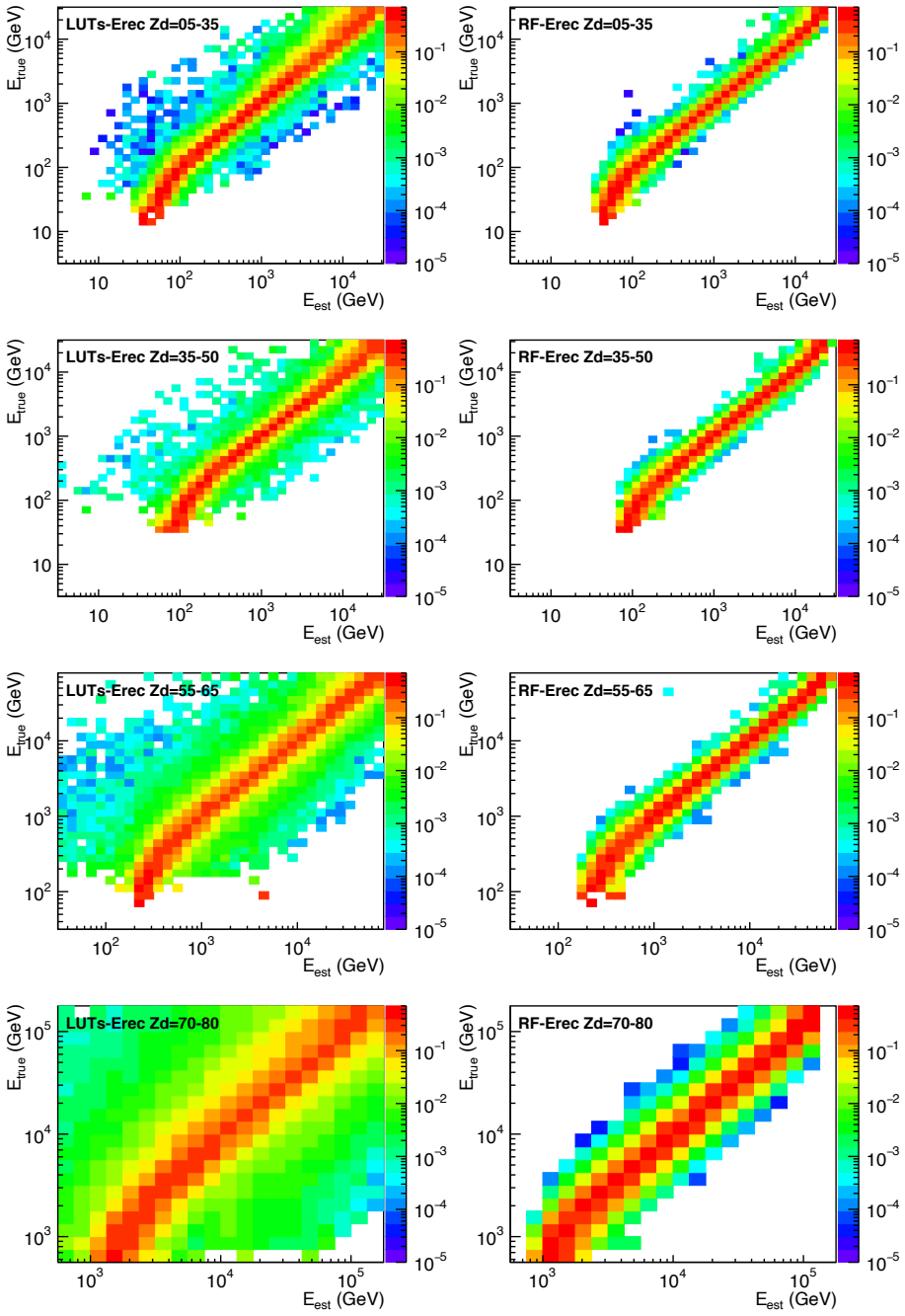


Figure 4.35: Migration matrices from LUTs-Erec and RF-Erec for various Zenith distance ranges

The migration matrices, derived with the LUTs-Erec (left) and RF-Erec (right), shown separately for various Zenith distance (Zd) ranges. The events are selected with the same size cut and efficiency cuts (see section 3.10.3), and the weight is applied based on the Zd distribution spent on the observation. The observation data used are the same as the ones for the validation through Crab Nebula spectrum, discussed in section 4.5.4.

4.6.2 The basic performance

Here I show the basic performance, namely the generalization error parametrised as the bias and resolution. The definition is explained in section 4.1.3. The performance, over wide range of energies, is reported in figure 4.36.

The result shows significant improvement of RF-Erec compared to LUTs-Erec in all Zd ranges. In case of low Zd range, the fitted energy resolution by RF-Erec reaches around 11 % above 1 TeV, while the one by LUTs-Erec never goes below $\sim$ 15 %, and becomes worse as energy increases, reaching $\sim$ 30 %. In the low energy side, the RF-Erec shows mostly $\sim$ 20% of fitted resolution, which is comparable to LUTs-Erec.

While the fitted resolution is to depict the tendency of majority population, the evaluation via standard deviation focuses on the tendency of outliers. To quantify the outliers, let us introduce tail spread.

$$TailSpread = (SD - \sigma)/\sigma \tag{4.6.1}$$

where SD is the standard deviation of the distribution, and σ is the fitted resolution, namely the width of the fitted Gaussian. The tail spread indicates the deviation from the Gauss distribution in the tail, giving large value for the larger numbers and/or the wider spread of the outliers[10]. The tail spread is shown in the bottom subpanels. The improvement is clear. For low Zd range the tail spread is reduced to less than half for entire energy range with stable bias. In the other Zd ranges, the improvement is more significant. Among the ranges shown here, Zd from 55 to 65 shows the biggest improvement, by reducing the tail spread to around 10 % of that with LUTs-Erec.

Even more, when the fitted resolution and the standard deviation are compared in higher Zd, the difference is even more dramatic. The discrepancy of standard deviation becomes always factor $\sim$ 2 or more, and the fitted resolution also approaches similar improvement as the Zd is larger.

RF-Erec shows better performance than LUTs-Erec also in bias. Bias by RF-Erec is always stable in the middle energy range, while LUTs-Erec shows instability in Zd from 35 to 50 and 70 to 80. Especially in Zd range from 70 to 80, the bias by LUTs-Erec cannot stay close to zero any more.

However, the bias always deviates from zero in the both edges. It can be naturally attributed to unbalanced energy distribution of train samples. As the estimated energy is determined in training by averaging the samples remaining at the last node after splitting, the population at the last node should ideally include equal amount with energy. For the energy estimator, I set the target value as logarithmic form, thus the distribution should be ideally uniform in logarithmic scale of energy, namely power-law distribution.

In highest energy side, the problem is the truncation of the distribution. The bias start to appear around the half of the maximum energy of the train data. This problem can be solved by extending the energy range of train data set.

In the lowest energy side, the deviation is larger. It would primarily come from the threshold effect. Since the lowest energy events survive the trigger and analysis only when it is bright enough, and resembles higher energy events, this threshold effect cannot be removed. Moreover there would be another

[10] For a Gauss distribution, the standard deviation is the width of the distribution. Since the fit focuses on the peak for a with Gaussian like distribution, the tail spread can be non-zero for the deviation from the Gauss fit in the tail.

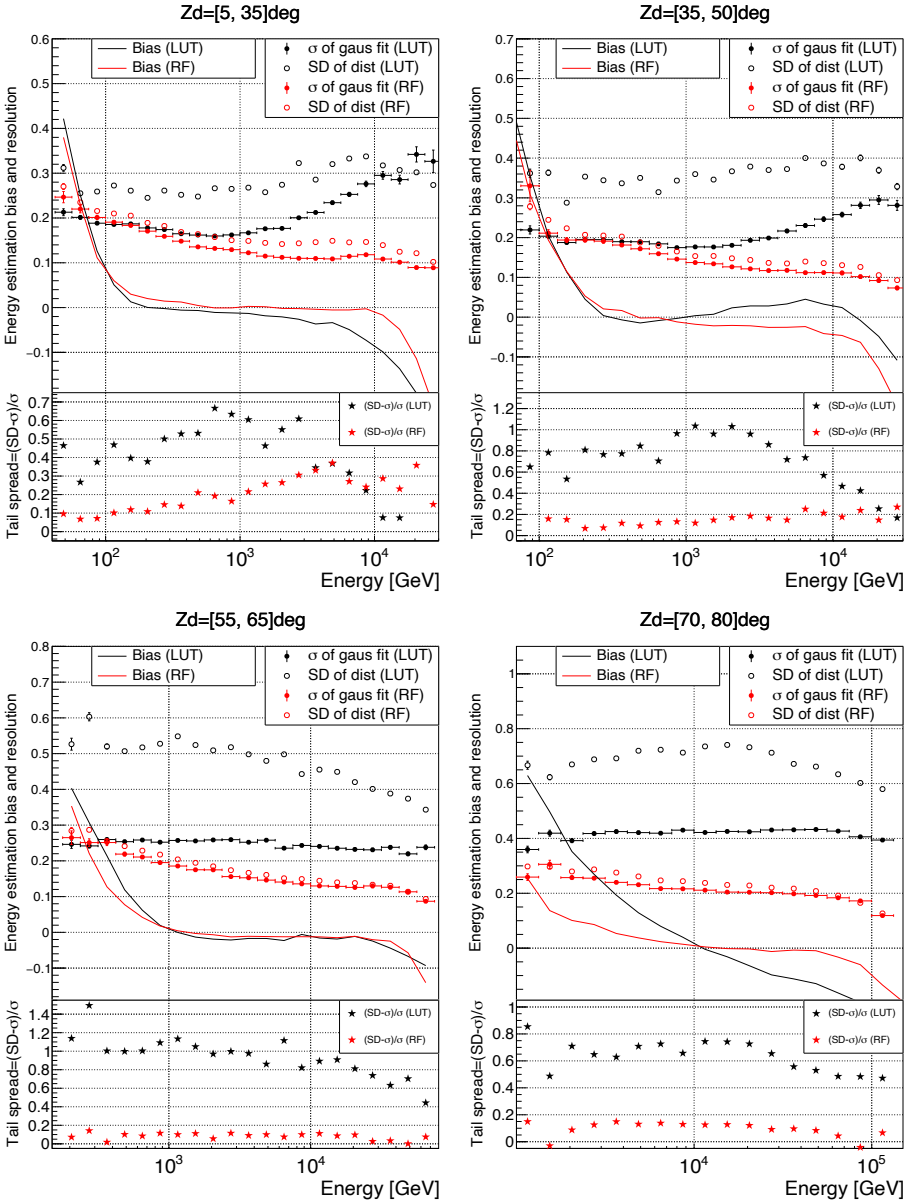


Figure 4.36: Basic performance of LUTs-Erec and RF-Erec in different Zd ranges

Performance of the LUTs-Erec (black) and RF-Erec (red) energy estimators are shown as a function of true energy, separately for four different ranges of Zenith distance (the ranges are shown on top of the plots).
In all panels, the top subpanels show the fitted bias (the solid lines), the fitted resolution derived from the Gaussian fit (the filled circles) and the resolution as standard deviation (SD) of the distributions of the quantity $(E_{est} - E_{true})/E_{true}$, where E_{est} is the reconstructed energy (the empty circles). For more details, see section 4.1.3. The bottom subpanels are the tail-spread, defined as the normalised deviation of the SD from the width of the fitted Gaussian.

reason to make the bias. The population of the training data set is distorted in energy because of the quick drop of the surviving probability of events in triggering and analysis. This distortion is critical to determine the prediction value because it is designed to meet the power-law distribution. The association between the energy bias and the energy distribution is noticeable, as I discuss in section E.2.2.

4.6.3 Bin reliability (purity and spillover)

As the migration matrix shows, the majority of the events are evaluated properly with RF-Erec; i.e. the E_{est} is similar to E_{true}. But even the small probability at the tail of the distribution can influence significantly the analysis. Since a spectrum of the γ-ray from the astrophysical source has power-law like distribution, and the collection area at high energy range is more or less flat, the number of signal events we have in an observation data will also follow power-law-like distribution. Therefore, the number of events in lower energy side are more abundant than in the higher energy side, and the tiny fraction of spillover to an estimated energy bin from the lower energy side results in a considerable amount. The steeper the spectrum, the larger the effect.

In this section I discuss bin purity and spillover effects. These evaluations focus on the reliability of the number of excess events in an histogram binned in energy, after statistical accumulation of misreconstruction.

Setup of the data generation

Figure 4.37 shows the schematic view of generating a histogram to evaluate. A spectrum of an imaginary γ-ray source is assumed, and it is convoluted with the telescope and analysis response. Since the response changes dependent on pointing position, a tracking history of observations with Crab Nebula was chosen, whose total observation time is approximately three hours. The collection area and migration matrix are generated on this tracking condition. The process takes into account two fluctuations. One is the Poissonian fluctuation on obtaining signals, and the other is on the energy migration, in which the estimated energy for each event is assigned based on the probability distribution for the correspondent true energy.

I evaluated two types of spectrum of the simulated source, in two types of Zenith distance observations[11]. The spectral types are continuous power law with typical index, -2.5, and steep index, -5. And the Zenith distance ranges are low Zd (from 5 to 35) and high Zd (from 55 to 65).

Bin purity

Bin purity is the fraction of excess events in a given estimated energy bin, whose true energy is also within the same energy range. We can judge how much a bin is free from contamination of the events with wrongly assigned energy.

In the migration process, the binning in estimated energy E_{est} and E_{true} are evaluated. For i-th bin, the bin purity is $N_{remain}(E^i_{true})/N(E^i_{est})$, where $N(E^i_{est})$ is the number of events ending up in the i-th bin of estimated energy,

[11] The observation times are 3.6 hours (3.7 hours) for the observations with Zenith distance range between 5 degrees and 35 degrees (55 degrees and 65 degrees).

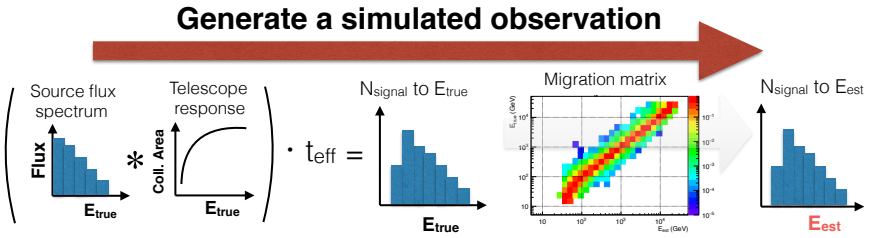


Figure 4.37: The scheme of generating a simulated observation for bin reliability evaluation

In this scheme, the simulated observation data takes into account only the simulated γ-rays, to focus on the effect of confusion in the energy reconstruction. To this end, the flux spectrum of γ-rays with effective observation time t_{eff} is given. The first step is to generate the distribution of number of γ-ray events (N_{signal},) over E_{true}, which is not accessible in the real data, and the second step is to convert it to that over E_{est}, which is accessible. In the first step, the flux is convoluted with the collection area, and multiplied by t_{eff} to obtain N_{signal}, then the Poisson fluctuation is given to N_{signal} in each E_{true} bin. In the second step, the conversion is performed using the migration matrix in a statistical way; the destination bin of E_{est} is determined statistically event by event, following the probability distribution in the migration matrix.

and $N_{remain}(E^i_{true})$ is the number of events to be delivered into the bin of the same energy range of E^i_{est} as that of E^i_{true}. The result is shown in figure 4.38[12]. I choose the binning to be 7 bins per decade in both E_{est} to that of E_{true}, so that a bin should contain sufficient quantile for the median energy event assuming a resolution of $\sim 15\%$.

The result shows significant difference between RF-Erec and LUTs-Erec. In the typical spectrum, while the purity of RF-Erec continues rising from 0.4 at lowest energy to 0.7 at highest energy, the LUTs-Erec never surpasses RF-Erec and start to drop after several hundred GeV. In the steep spectrum, LUTs-Erec performs always worse, and the difference is larger. Both RF-Erec and LUTs-Erec show lowering of the bin purity in the low energy side. Not only the worsening of the energy resolution in the lowest energy range, but also the energy range below hundreds of GeV suffers from the rising bias, and the events start to escape from the correct energy bin to higher energy bin. Moreover, small spillover from lowest energy bin contributes large fraction to the higher energy bin because of the power-law energy distribution.

For the high Zd observation, the difference is more dramatic. The steep spectrum comparison in HZd shows that the purity by LUTs-Erec is below 2 % above 2 TeV in LUTs-Erec, while that by RF-Erec is above 40%.

Spillover

As part of spillover (the migration of events to wrong energy bin), I quantify the (fake) excess events in a energy bin where no excess event is expected. Let

[12] For the steep source, the evaluation was done separately for low energy and high energy, because the number of events are enormously different. The low energy and high energy evaluation has the overlap region so that the purity continues smoothly.

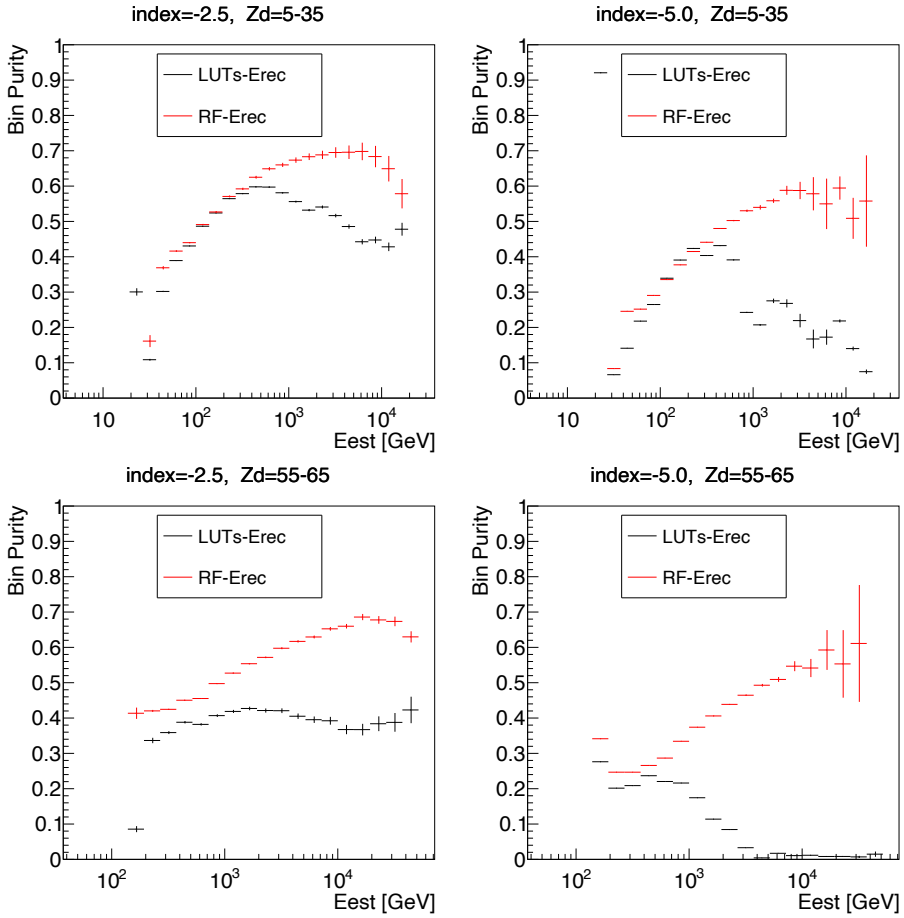


Figure 4.38: Bin purity vs Energy with LUTs-Erec and RF-Erec

The bin purities as a function of energy, with LUTs-Erec and RF-Erec, are shown for two power-law indices and two Zd ranges, denoted in the panels.

us consider a spectrum of true energy E_{true}, as power law with a hard cutoff at energy $E_{\rm cut}$, to see clearly the effect of the spillover. A wide migration matrix changes the observed spectrum over estimated energy E_{est} to smear out the intrinsic cutoff, and events with E_{est} beyond $E_{\rm cut}$ appear. Having the excess events, an analyzer is forced to consider the possibility of observing events with E_{true} beyond $E_{\rm cut}$. In this way, a wide migration matrix adds confusion to the investigation of the highest energy observed, from a γ-ray source, which can lead to profound physics implications.

To this end, I set a hard cut off at energy of 1 TeV in the aforementioned power-law spectra of a simulated γ-ray source. After the generation of a simulated observation, the histograms of E_{true} and E_{est} are compared bin by bin. The figure 4.39 shows the ratio of the number of events in E_{est}, denoted as $N(E_{est})$, to that in E_{true}, denoted as $N(E_{true})$. Above the cutoff energy of 1 TeV, there must be no event in E_{true}. Thus the denominator is set to the num-

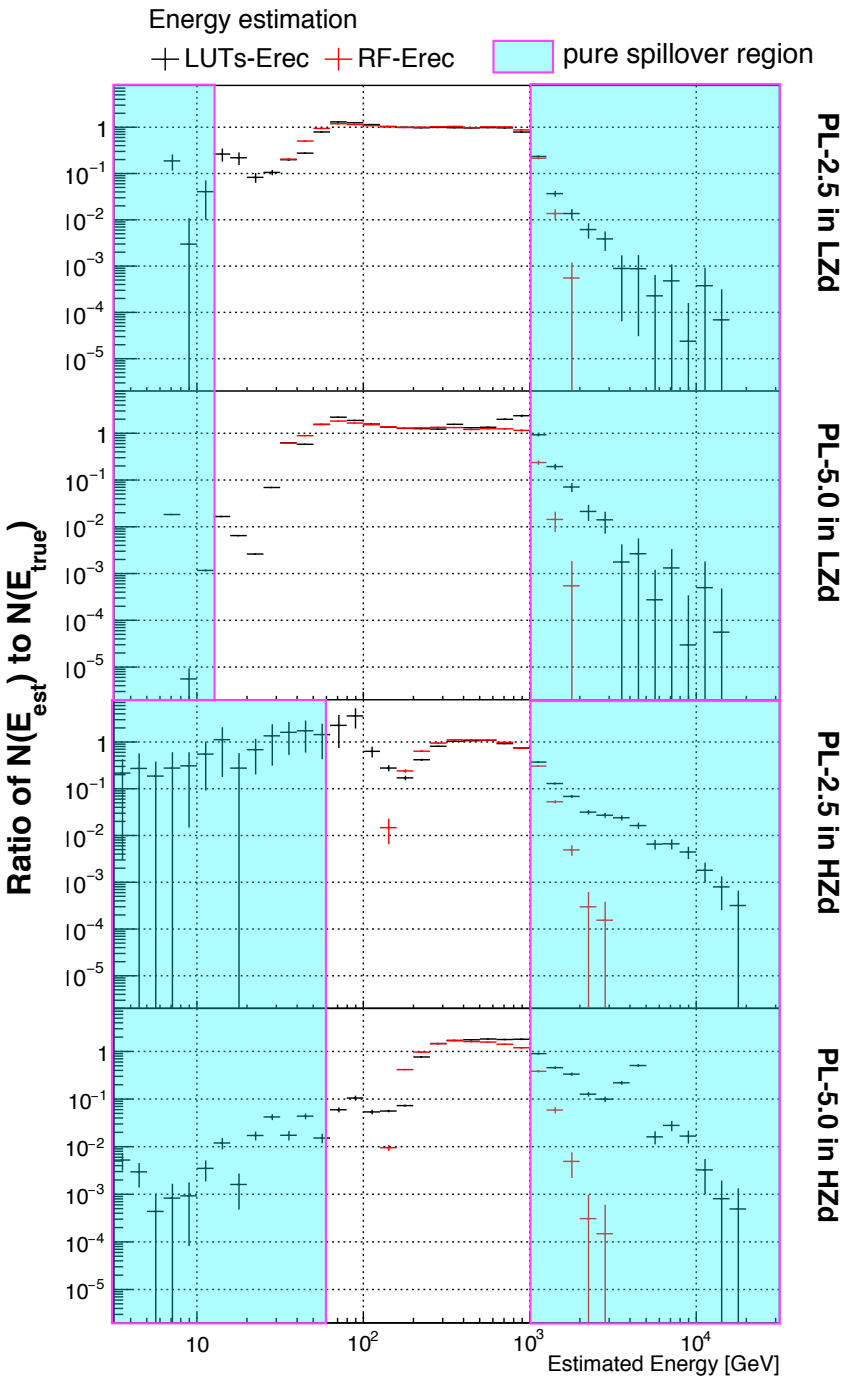


Figure 4.39: Spillover ratio vs energy with LUTs-Erec and RF-Erec, for spectra with hard cutoff at 1 TeV

The spillover ratio, which is the ratio of the number of events of E_{est} in the same bin of E_{true} to that of E_{true} is plotted. Two power-law spectra with index -2.5 and -5 (indicated in the right of the panels) with a hard cutoff at 1 TeV, are investigated in the two Zd ranges of observation. Due to the hard cutoff, there is no event above 1 TeV of true energy (pure spillover region), thus the spillover ratio is computed as the ratio to the number of events in the last bin below 1 TeV. Likewise, the lowest energy is also pure spillover region, there the ratio is computed as the ratio to the number of events in the lowest energy bin of true energy.

ber of events (in E_{true}) at the last bin below 1 TeV. The lowest energy side also has the bins with simulated excess events, and they are treated similarly[13]. The number of bins in E_{true} and E_{est} are set to be equivalent as 10 per decade to see the spillover tendency in detail. I discuss the result in the following, separately for below and above 1 TeV.

As the main property below 1 TeV, the ratio should ideally be 1. For the typical spectrum in Low Zd, the ratio is close to 1 for both RF-Erec and LUTs-Erec. It shifts positively in steep spectrum or/and in High Zd, because of the power-law spectrum. The distribution of numbers of events is not balanced over energy, so that the number of events coming in from lower energy bins is always higher than that going out from higher energy bins. The steeper the slope of power law or the higher Zd, the more the spillover, resulting in the higher ratio.

Below 1 TeV, the differences between RF-Erec and LUTs-Erec lie in the ranges close to 1 TeV and lowest energies.

In the range close to 1 TeV, the difference is significant in the steep spectrum. Despite the higher spillover due to the steep spectrum, RF-Erec lets the ratio approach 1 for higher energy by better energy resolution. However the ratio by LUTs-Erec does not improve for higher energy, and it shows even quick rise to above 2 in the case with steep spectrum in low Zd. It indicates that the bins are already dominated by the contamination from different energies of E_{true}, and it is consistent with the drop of bin purity shown in figure 4.38. Therefore RF-Erec improves the bin reliability especially in the high energy range.

In lowest energies, LUTs-Erec shows the long tails in all the cases, although the expectation is 0 below energy threshold. The ratio is higher in the typical index case. This tells the large number of events in lowest energies spillover into even lower energies in LUTs-Erec, and such escape of events can clearly be seen by lower ratios in the energies of several tens of GeV in low Zd and a few hundreds of GeV in high Zd. Therefore the RF-Erec can lower the energy threshold, because the higher ratio, kept in the lowest energies, results in higher number of excess events.

As for the energy range above the cutoff energy of 1 TeV, the value should be 0, as there is no signal. However the spillover brings the (fake) excess events there. In all the settings, RF-Erec shows good performance; the ratio quickly falls down above 1 TeV and disappears already at 2 TeV, on the other hand, LUTs-Erec shows a long tail continuing to even above 10 TeV.

And in the region above 1 TeV, the systematic effect can be seen in the result by LUTs-Erec for the steep spectrum in high Zd. The distribution above 1 TeV is not just a decreasing tail, but is wiggled. The reason is presumably due to the wide distribution of the migration probability, together with low statistics of the MC data for the migration matrix. The range of estimated energy for a true energy is wide for LUTs-Erec, however the number of events to be used for constructing the migration matrix is low in the low energy range. Therefore the probability distribution in the migration matrix becomes sparse and wiggled, which is of course different from the real probability distribution of migration. In the generation of the simulated observation data, the distribution is generated using the migration matrix and the spillover from the low energy range into the high energy range is enhanced by very high flux.

[13] The lowest energy bin of continuous positive excess bins is chosen as the reference bin. Below the bin, the ratio of the excess events of E_{est} in each bin to that of E_{true} in the reference bin is calculated.

4.6.4 Effects on unfolding

Here I show a direct influence of the energy estimation performance on modeling the spectrum of a source.

As discussed in section 3.11, the accuracy of energy reconstruction is always limited, and the energy distribution of detected events is deformed from what it is in the real energy. Therefore there must be spectral unfolding, a process to take into account the deformation when the energy spectrum of the flux is reconstructed. And the performance of energy reconstruction will appear in this process, where a better energy reconstruction results in smaller spectral deformation, hence better spectral unfolding. We deal with such inverse problem by forward folding or unfolding. In case of forward folding, we must assume a model of the initial spectrum to explain the observed distribution as a function of the estimated energy. On the other hand, unfolding does not require any model, however it instead imposes a regularization in the unfolded distribution like a smoothness of the spectrum to force the solution to converge. Both methods have pros and cons, and they work complementary to each other. Thus the most important point for practical use of unfolding methods would be to obtain consistent results via both methods.

This subsection focuses on evaluating the consistency between the true spectrum and the one reconstructed. To this end, I generate a set of simulated observation data with given spectra, to be compared with the results of analysis on them. I adopt a very bright source with very steep spectrum for the simulated observation data, to be an simple and clear example.

Setup of the data generation

Figure 4.40 shows the schematic view of the strategy used to generate the histograms to evaluate. In addition to signal generation process, the background events are generated. The "ON" distribution is the sum of signal distribution and background distribution. And the three "OFF" regions all consist of background events. these background events are taken from the real data[14], giving Poissonian fluctuation. The OFF distribution is the average of the three OFF regions. Then the analysis can determine the number of signal events only by the excess events, namely the subtraction of OFF distribution from ON distribution.

In this evaluation, two migration matrices with different binning are used to take into account the energy migration more in detail. One is 280 bins in true energy and 28 bins in estimated energy. This tries to simulate more detailed migration process in generating the signal for the simulated observation data. The other is 20 bins in true energy and 28 bins in estimated energy, to be used for normal unfolding[15].

I consider a source with a power-law index -5 and the flux level of 10 C.U. at 500 GeV. The observation condition is set to be 3 hours of total observation time and high Zd range. The data analysis of unfolding needs both "ON" distribution and "OFF" distribution. The number of OFF regions is set to be 3. They are summarised in table 4.4.

[14] The background distribution depends on the tracking history, because it varies with Zenith distance and Azimuthal angle.

[15] To make the unfolding process converge, coarse binning in true energy is normally adopted.

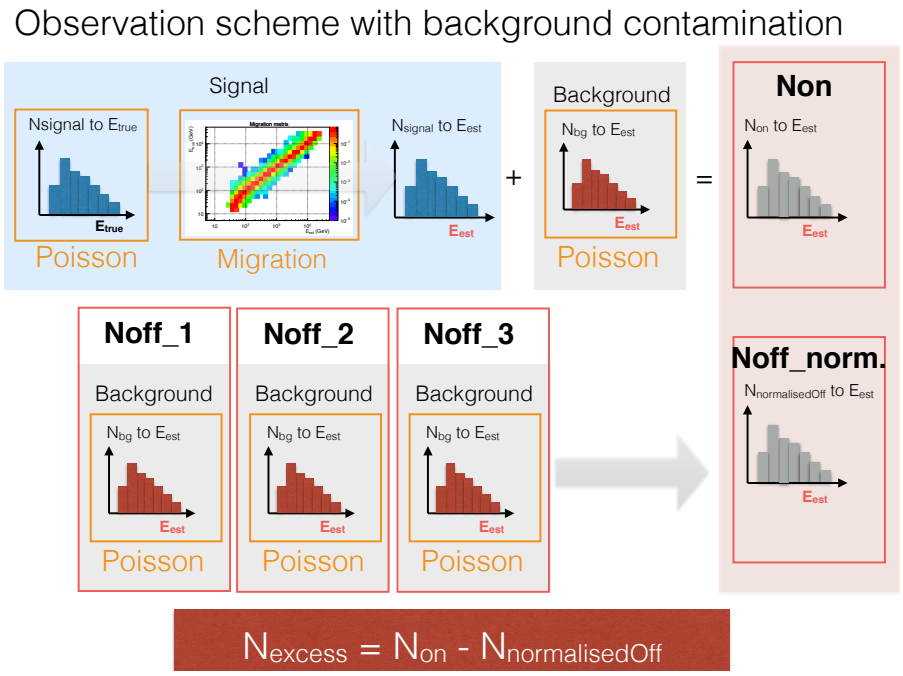


Figure 4.40: The scheme of generating a simulated observation for unfolding performance evaluation

In spectral unfolding, the distributions to be given to the analysis are, the number of events in ON region and in OFF regions, as a function of estimated energy E_{est}. The number of excess events, $N_{\mathrm{ON}} - N_{\mathrm{normalisedOFF}}$, is used for unfolding, and the number of ON and OFF events, N_{ON}, N_{OFF}, are used in forward folding.

To this end, the schematic view shows the generation of signal events (in the blue region) and the background events (in the gray regions) for the ON region and the OFF regions. The signal events is generated in the same way as figure 4.37 (the convolution step with collection area is omitted). The distributions of background events are calculated from the template distribution by adding Poissonian fluctuation. The template distribution is obtained from the real observation data of the Crab Nebula. In the schematic views, all the statistical fluctuations applied are written in orange texts.

Table 4.4: The condition parameters for unfolding evaluation

	Parameter	Value
Source	Slope	-5
	Amplitude	$5.6 \cdot 10^{-11}$
Observation and anaylsis	Zd range	55 - 65
	Observation time	3 h
	Number of OFFs	3

The source parameters are to formulate the differential flux of γ-ray, $dF/dE = A \cdot \left(\frac{E}{1\mathrm{TeV}}\right)^{\alpha}$ [$\mathrm{TeV}^{-1}\cdot\mathrm{cm}^{-2}\cdot\mathrm{s}^{-1}$] , where α is slope, A is Amplitude and E is energy. By the parameters above, the flux level is adjusted to be 10 C.U. at 500 GeV.

Setup of the analysis

The forward folding does not require any special treatment, because it requires only the model spectral shape. The model is fitted to explain the measured distributions of both ON and OFF regions, also taking into account the fluctuation of signal and background events.

On the other hand, unfolding aims at no strict assumption on the spectral shape. Instead it needs more elaborated configuration in the bin selection of estimated energy and true energy bins. The configuration should be considered more carefully than adopting the standard selection, because the spectral shape is unusual like the case I evaluate here. It is challenging to recover a steep spectrum because the energy distribution of the excess events is heavily distorted by the confusion in energy estimation. A tiny fraction of events spilling over to another energy bin sometimes becomes majority in the destination bin in estimated energy, as shown in section 4.6.3. Therefore the unfolding needs to consider as many bins as possible to reduce systematic effect, as long as the contribution from the background events is small.

To this end, the bin range of estimated energy is selected in which all the consecutive bins have at least 4 excess events. Typically, the minimum number of excess events is 10, but this limit would be too tight in steep spectrum case, and it would deform the unfolded spectral shape at the edge. Therefore, for this study I set the minimum number of excess as 4.

On the other hand, the criteria to choose the true energy bins is based on collection area and migration matrix. In the low energy side, the edge is determined by the collection area because the bins below energy threshold do not provide many events. In the formal limit the minimum collection area of the bin to be selected is 1000 m^2. However a steep spectrum generates significant amount of events despite the small collection area, thus I chose the low energy side bins as long as the collection area is non-zero. For the high energy side, I chose the criteria in the same way as the typical one; the true energy bins whose sum of migration probability covered by the selected estimated energy bins is more than 70 %.

The actual selected bin ranges are shown in figure 4.41.

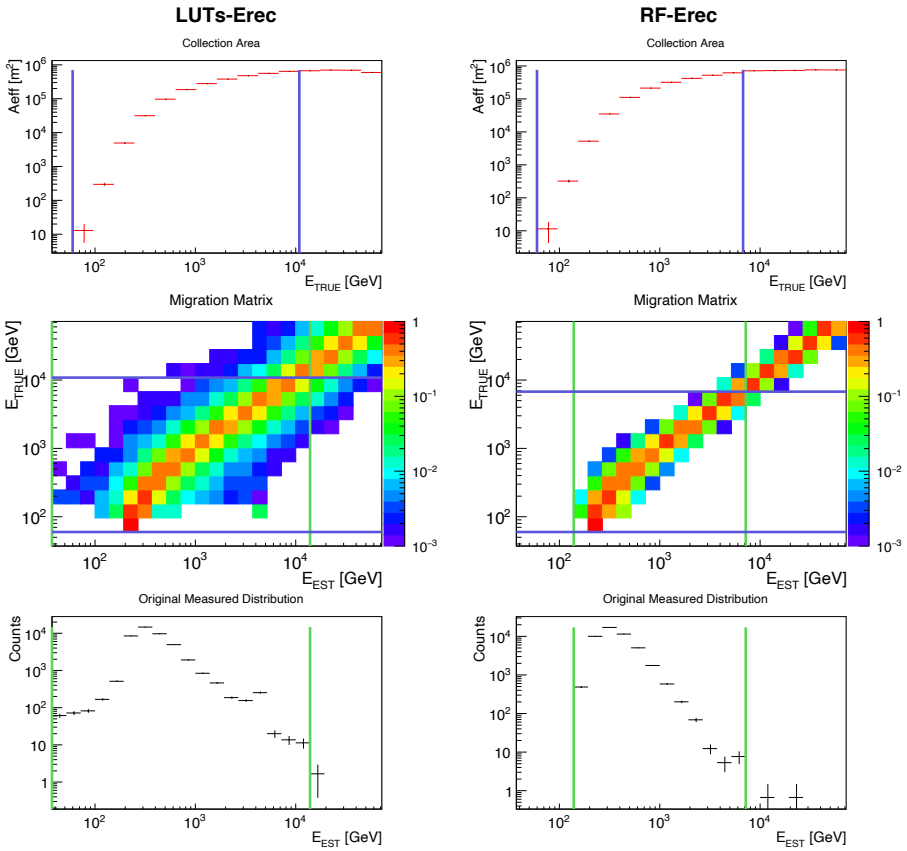


Figure 4.41: Collection area, migration matrix and excess events of simulated observation data with LUTs-Erec and RF-Erec

The ranges for calculating the unfolding are also shown as the colored bars in the plots. The green (blue) bars are the lowest and highest edge of the estimated (true) energy range.

The result of forward folding

The forward folding is a fit of a model function to the measured distribution with maximum likelihood process. The result comes as the fitted parameters of model spectrum, together with the goodness of the fit. In this performance evaluation, the model spectrum is set to be a power law.

$$\frac{dF}{dE} = A_0 \left(\frac{E}{E_0} \right)^{\Gamma_{ph}} \tag{4.6.2}$$

where $\frac{dF}{dE}$ [$\mathrm{TeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$] is the differential flux as a function of energy E, A_0 is the amplitude, Γ_{ph} is spectral index, and E_0=1 TeV is the pivot energy. The parameters obtained and the goodness of fit are shown in table 4.5.

Both RF-Erec and LUTs-Erec show good agreement with the true spectrum. The model parameters are estimated to be consistent with the original spectrum

Table 4.5: Forward folding results for the simulated observation data

	A_0 [10^{-11}/TeV cm^2 s]	Γ_{ph}	χ^2/NDF
LUTs-Erec	5.79 ± 0.31	-4.99 ± 0.05	6.3/11
RF-Erec	5.68 ± 0.15	-5.00 ± 0.03	8.4/11

shown in table 4.4 at almost the same precision, and there is no clear difference of energy reconstruction performance between them, although RF-Erec shows closer estimated power-law parameters to the original spectrum with smaller uncertainty. However the difference can be understood deeper in the control plots and these contain major differences to possibly affect the result in real cases.

In the control plots shown in figure 4.42, the top two panels compare the measured values with the expected ones based on the spectral model with the maximum likelihood. Both strategies show good agreements between the expected and observed distributions both in ON and OFF regions. The agreement can also be assessed via the residual and the χ^2 contribution plots, shown in the 3rd and 4th rows of the figure. However, the distributions are different between the two strategies especially in ON events, where an enhancement at about 4 TeV can be seen in the ON distribution with LUTs-Erec (it can also be seen in the excess events distribution shown in figure 4.41).

The reason is identified via the assessment on the collection area and the migration matrix, shown in figure 4.41. There is no anomaly in the collection area, thus this is not due to the enhancement of the sensitivity at the energies. Instead it is from the migration probability from low true energy to the high energy, as the migration matrix for LUTs-Erec shows an isolated high probability bin at about 100 GeV of true energy vs about 4 TeV of estimated energy. This small probability of spillover is enhanced by the steep power-law spectrum, and gives the large impact in the destination estimated energy. This problem is also discussed in section 4.6.3 and section E.2.1.

Although the distribution has such anomaly, the forward folding performs well, because it already "knows" the anomaly; both of the migration matrices, namely the one for performing migration in producing the simulated observation data and the other for spectral unfolding, have the same distributions of the migration probabilities although the different binnings are used, because they are generated from the same test samples.

Therefore, if different sets of simulation data were provided for the simulated observation data and the spectral unfolding, the forward folding with LUTs-Erec would have performed worse than that with RF-Erec, because the wider distribution and isolated bins in the migration matrix can make larger difference of distribution patterns when different sets of test samples are used. This indicates that the forward folding with LUTs-Erec suffers from larger systematic uncertainty in the analysis also for a real observation data, because the migration matrix used for the spectral unfolding is different from the real migration process.

The bottom panels of the control plots show the contamination fraction, which is the fraction of events in an estimated energy bin migrated from the energies higher or lower by more than 20 %. The contamination fraction with

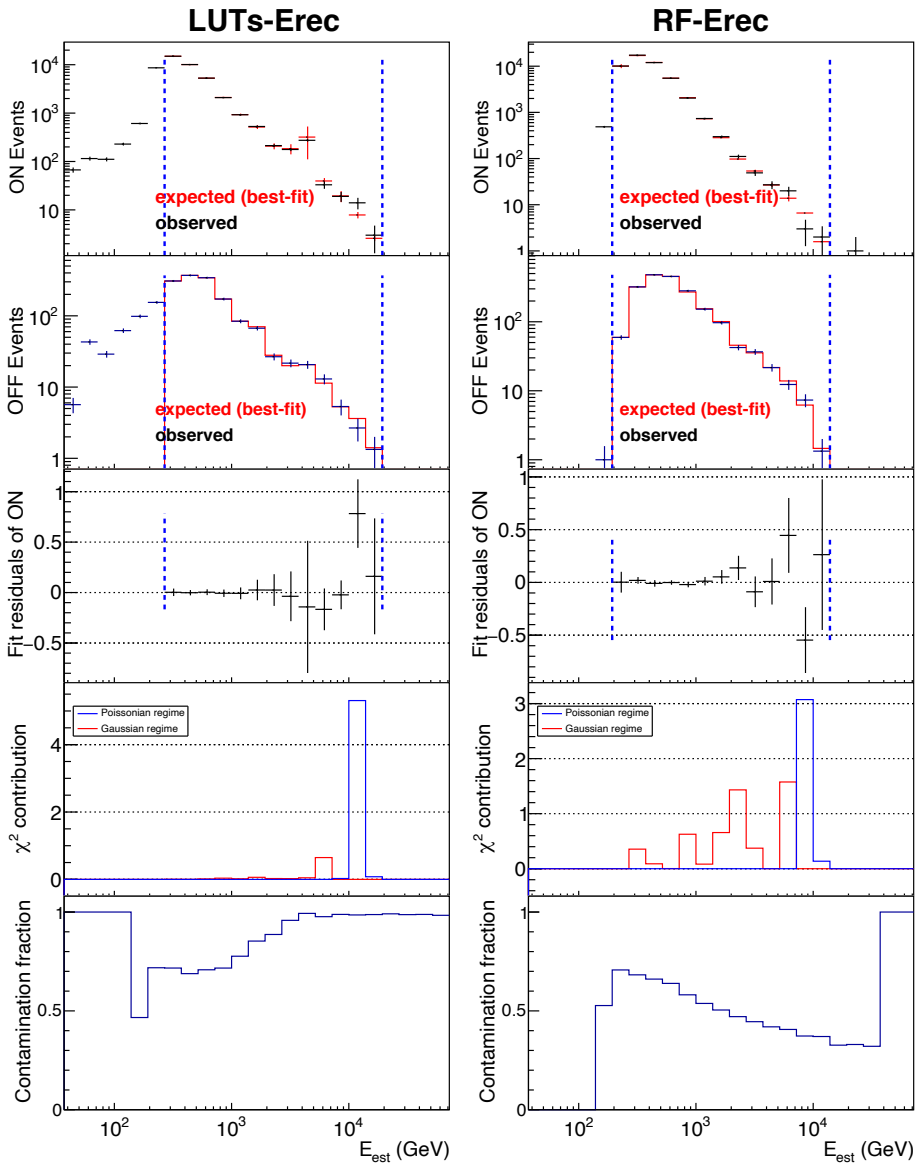


Figure 4.42: Control plots from forward folding with LUTs-Erec and RF-Erec for simulated observation data

The consistency of the result of forward folding is shown in the different point of views. The top two rows are the number of events in ON and OFF regions, where the reconstructed number of events from the fitted model spectrum (red markers) is shown together with the observed number of events (black markers). The vertical dashed lines indicate the range to calculate the likelihood for fitting. The third row is the residual of the reconstructed number of events with respect to the measured number of events. The fourth row is the χ^2 contribution. The bottom row is the contamination fraction, which is the fraction of the spillover events migrated from the energies higher or lower by more than 20 %.

LUTs-Erec is always high from about 0.7 below 1 TeV and even higher above 1 TeV, reaching almost 1 at several TeV. The one with RF-Erec is lower for higher energies, reaching below 0.4 at several TeV. As discussed in section 4.6.3, the high contamination fraction drops the bin purity and it affects the flux reconstruction for the individual energy bins based on the forward folding.

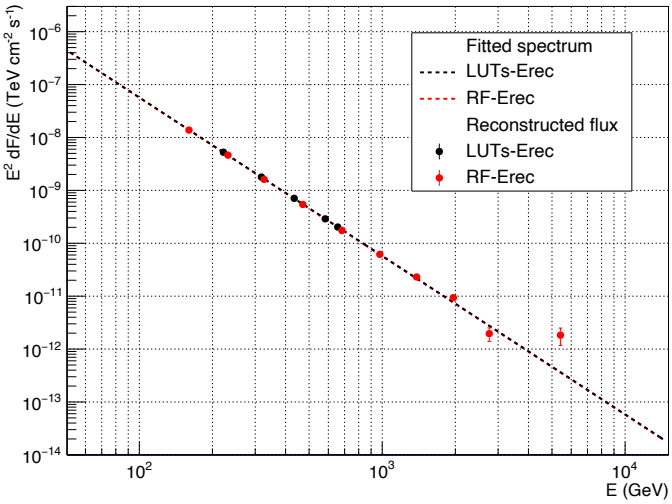


Figure 4.43: The spectra of simulated observation data, reconstructed by forward folding with LUTs-Erec and RF-Erec

The reconstructed spectra by LUTs-Erec and RF-Erec are drawn overlaid in different colors, shown in the legend. The lines are the fitted model spectra, and the points are the reconstructed flux, using the spectra for taking into account the energy migration in the collection area. The parameters of the true spectrum are shown in table 4.4, and the ones obtained by forward folding and their goodness of fit are shown in table 4.5.

Figure 4.43 shows the reconstructed spectral points via forward folding, together with the fitted model spectra drawn as dashed lines. The flux points are calculated from the number of excess events, the observation time and the collection area, but the collection area takes into account the spectral shape and the energy migration. For this reason, the points are plotted at the median true energy (according to MC) of the γ-ray events from which they are calculated. If the contamination fraction becomes higher, the median true energy in a bin becomes far off the corresponding energy range, thus such data points are removed.

Consequently, there are differences in the number of data points and the positions of them, between RF-Erec and LUTs-Erec. The difference clearly shows that RF-Erec significantly improves the reliability of data points by enabling the estimation of the flux data points above 1 TeV.

The result of unfolding

The spectral points reconstructed by unfolding are shown in figure 4.44. There are different regularization methods as described in section 3.11.2, and are in-

dicated by color. The black points overlaid are the reconstructed points by forward folding.

The consistency of these reconstructed spectral points with the measured excess event distributions is shown in table 4.6, together with the results of power-law fit to the spectra.

Table 4.6: Results from unfolding and power-law fits to unfolded data points of simulated observation data with LUTs-Erec and RF-Erec

(a) LUTs-Erec

	Unfolding	**Power-law fit**		
	χ^2_0/n_y	A_0	Γ_{ph}	χ^2_{PL}/NDF
Tikhonov	3.7 / 18	5.91 ± 0.19	-4.96 ± 0.03	8.9 / 7
Schmelling	99.9 / 18	7.22 ± 0.16	-4.74 ± 0.02	53.0 / 9
Schmelling(m)	674.3 / 18	1.49 ± 0.02	-3.81 ± 0.02	635.6 / 9
Bertero	47.9 / 18	7.61 ± 0.1	-4.71 ± 0.02	82.1 / 6
Bertero(w)	179.3 / 18	1.01 ± 0.02	-4.32 ± 0.02	239.8 / 8

(b) RF-Erec

	Unfolding	**Power-law fit**		
	χ^2_0/n_y	A_0	Γ_{ph}	χ^2_{PL}/NDF
Tikhonov	2.8 / 12	5.72 ± 0.10	-4.97 ± 0.02	29.6 / 8
Schmelling	223.1 / 12	5.88 ± 0.12	-4.90 ± 0.02	34.9 / 8
Schmelling(m)	725.4 / 12	3.27 ± 0.08	-5.09 ± 0.03	295.4 / 7
Bertero	27.5 / 12	6.00 ± 0.13	-4.93 ± 0.02	59.6 / 8
Bertero(w)	34.7 / 12	6.72 ± 0.13	-4.79 ± 0.02	159.0 / 7

The left column shows the consistency of the unfolding result, χ^2_0/n_y, where χ^2_0 is the chi-square between back-folded distribution of number of excess events and the one measured, and n_y is the number of bins of estimated energy which were taken into account in unfolding process. The right columns are the result of the power-law fit to the unfolded spectra, performed with the correspondent regularization methods, using RF-Erec and LUTs-Erec. The power-law function used for the fit is $\frac{dF}{dE} = A_0 \left(\frac{E}{E_0}\right)^{\Gamma_{ph}}$, where $\frac{dF}{dE}$ [$\mathrm{TeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$] is the differential flux as a function of energy E, A_0 is the amplitude, shown in the unit of [$10^{-11}/\mathrm{TeV}\,\mathrm{cm}^2\,\mathrm{s}$] in the table, Γ_{ph} is the spectral index, and E_0 is the pivot energy set to be 1 TeV.

It is not sufficient to compare just the data points drawn here. In each regularization strategy, there are five different criteria for choosing the best strength of the regularization, explained in section H.1. The results shown here are only by the point of comparable noise levels between before and after unfolding, as the representative criterion. Moreover the results are not obvious for comparing the performance of energy reconstruction. In addition to the data points,

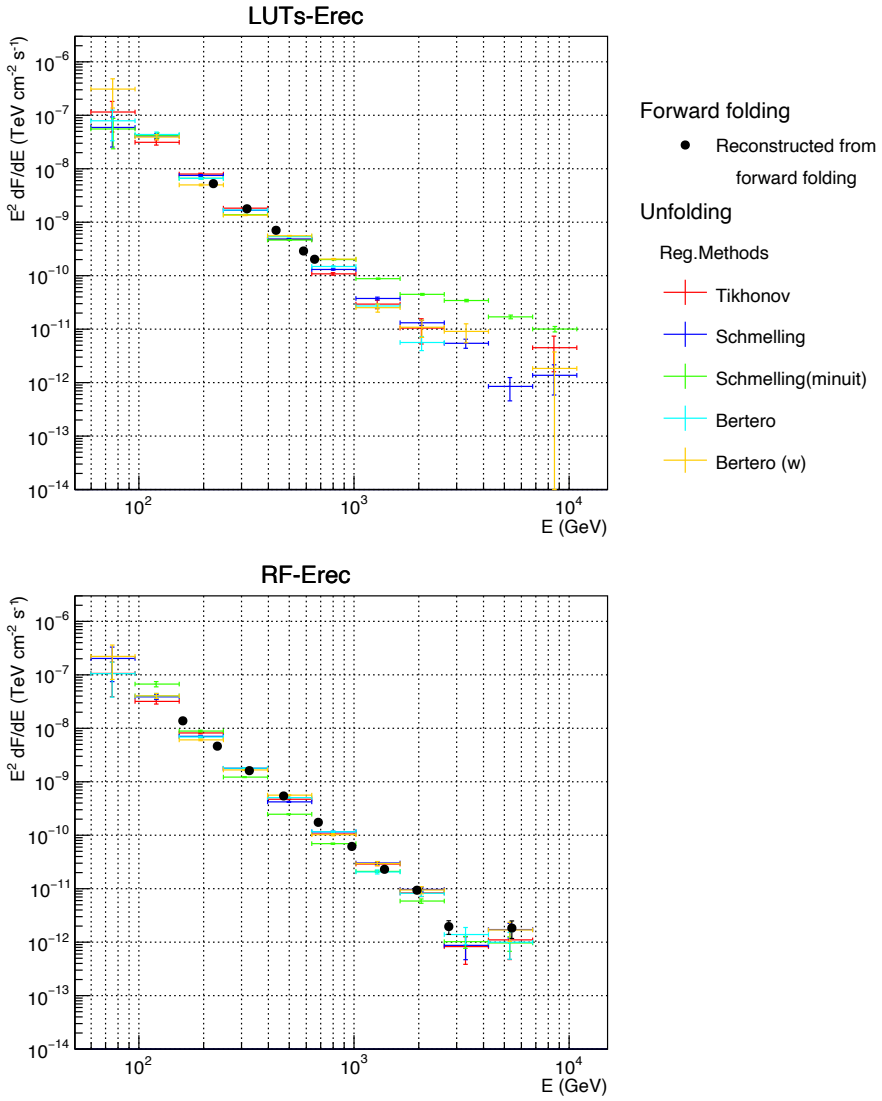


Figure 4.44: The spectra of simulated observation data, reconstructed by unfolding with LUTs-Erec and RF-Erec

Simulated observation of a source with a power-law index -5 and the flux level of 10 C.U. at 500 GeV, whose observation parameters are denoted in table 4.4. The results are separately drawn for the ones with the LUTs-Erec and RF-Erec. The colored points are the reconstructed flux by unfolding with different regularization methods, shown in the legend. The black points show the reconstructed flux based on forward folding.

the consistency in many aspects should be satisfied. The main idea on how to examine the results is summarized in section H.2. Since the evaluation is complicated, I begin with discussing the results on the representative criterion, and continue to the detailed comparison among the different criteria for the best regularization strength, together with the different regularization methods.

Comparing the reconstructed spectral data points by the representative criterion, figure 4.44, the results with RF-Erec are consistent among all of the regularization methods and they are consistent also with the data points by forward folding. On the other hand, the unfolding with LUTs-Erec shows different results among the different regularization methods, also showing deviation from the reconstructed spectral line by forward folding above 1 TeV, especially in the Schmelling(minuit) method. Moreover, some regularization methods show the missing data points in the middle before the highest energy point. This is not physically reasonable thus it should be avoided.

The reason behind this can be clarified by the control plots shown in figure H.3, together with the detailed discussion. Here I summarize the difference in unfolding between RF-Erec and LUTs-Erec, that resulted from the energy reconstruction performance. All the differences indicate that RF-Erec performs better.

- LUTs-Erec shows larger deviations from the true spectrum

  The unfolding with LUTs-Erec in general results in worse reconstruction than that with RF-Erec. While the true spectrum is a simple power law with the index of -5, the power-law fit to the reconstructed points are typically significantly harder with LUTs-Erec, as shown in table 4.6. The χ^2_{PL} for the fit are also worse, even after comparing among the different criteria for the best regularization strength.

  Moreover, the unfolded spectrum with LUTs-Erec sometimes indicates anomaly with respect to a power-law shape. There are sometimes a discontinuity in the data points above several TeV, cutoff or an upturn towards high energy above 1 TeV. The presence of such artificial anomalies would challenge the scientific interpretation of the results.

- LUTs-Erec shows inconsistent results among different weight choices

  In the real observation, the real spectrum is not known, thus the different results among different ways of unfolding would complicate the interpretation of the results. While RF-Erec shows mostly consistent results among all, LUTs-Erec gives different results, not only among the different regularization methods, but also among the different criteria for the best regularization strength within the same regularization.

- LUTs-Erec shows problems to converge

  Unfolding with LUTs-Erec sometimes shows a failure in convergence for finding a χ^2 minimum. It can still determine the solution because it is chosen from the χ^2 minima for different strengths of regularization, however it is less reliable for the choice of the best weight.

Chapter 5

RF-Erec enables the spectral analysis of the first Gamma Ray Burst detected at TeV energies

On Jan 14 in 2019, MAGIC detected the clear signal from a γ-ray burst (GRB), named GRB190114C. This was the first time a GRB was significantly detected at VHE γ-rays. The significance of the detection from the source at redshift 0.4245 was 50 sigma, and the estimated energies of the thousands of photons ranged from 0.2 TeV to around 1 TeV.

Due to such cosmological distance, they are strongly affected by Extragalactic Background Light (EBL) absorption, and the absorption becomes very strong for the highest energy photons. The TeV photons from GRB190114C are the most distant ones detected to date, and hence the ones suffering the strongest EBL attenuation. This means that the spectrum becomes so soft that it becomes challenging to reconstruct. If there had not been the novel method available, the spectral analysis of these data with LUTs-Erec would have led to large biases in the final results, as I show in section 4.6 and later on in this chapter. Fortunately, I had provided the novel energy reconstruction using the RF-Erec just on time, and the spectral analysis of these data was successfully conducted, yielding two scientific publications in Nature about the physics of GRBs [8, 111], and one on PRL about constraints on Lorentz Invariance Violation [83].

However, the latest version of my energy reconstruction (RF-Erec-v5) was not yet available by the time of preparation of these papers, and an intermediate version (RF-Erec-v3) had to be used. As introduced in section 4.4.5, RF-Erec-v3 had some issues that may introduce small bias in the reconstructed spectra, but it was already superior than LUTs-Erec, and hence the MAGIC collaboration decided to use it for the spectral analyses of this unprecedented data set.

In this chapter I show the better analysis using RF-Erec-v5 energy reconstruction and the alternative cut condition determined after further investigation on the observation and simulation data. Although the major tendencies of spectral shape are consistent,i.e., the main physical properties do not change

from the publication, my analysis show higher statistics and thus more information at higher energy. In addition, I will also present the spectral analysis using the LUTs-Erec, which is the result we would have obtained if my energy estimation method had not been available.

This chapter composes as follows. In section 5.1, I introduce GRB, focusing on the measure issues related to the VHE γ-rays. The observation of the GRB190114C is summarised in section 5.2 describing the outline of observations, the basic properties indicated from the analysis of the observations, and the analysis of the MAGIC observation data. The analysis method is described in section 5.3. The data taking condition was atypical, and the data needed additional inspection to adjust the analysis. In section 5.4, I compare the analyses between the latest RF energy estimation (RF-Erec-v5) with LUTs-Erec. The comparison of my analysis with the standard GRB analysis is discussed in appendix I. In section 5.5, I summarise the interpretation of the data, featuring the TeV component of the emission from GRB190114C observed with the MAGIC telescopes.

5.1 Introduction to Gamma-Ray Bursts

Gamma-Ray Bursts (GRBs) are a burst of γ-rays. To more specific, it is a sudden release of emission that lasts from milliseconds to thousands of seconds, in the energy range from tens of keV to several MeV, which is the typical bandpass of spaceborne GRB detectors [117]. The burst accompanies the emission in the other energies of electromagnetic wave, as well as much longer and different temporal profiles.

Physically, GRBs are the most luminous explosions in the Universe, although their locations are cosmologially distant even up to redshift 10. The typical isotropic γ-ray luminosity is $\sim 10^{51} - 10^{53}$ erg s^{-1}. Compared to the Sun, with $\sim 10^{33}$ erg s^{-1}, this energy release corresponds to the energy released by the Sun in its entire 10-billion-year lifetime. And this value is far above even the emission from an Active Galactic Nuclei (AGNs), which is the central active region of a galaxy, dwarfed by a GRB. Powered by accreting super-massive black holes, the most energetic AGN luminosity is merely $\sim 10^{48}$ erg s^{-1}.

Despite the enormous energy release, the variability seen in the explosion is very high with the timescale of down to ms. This apparent compactness problem requires that the source emitting the γ-rays of GRB must move relativistically towards Earth [96]. The boost effect with relativistic speed lowers the photon energy at the comoving frame of production site and enlarges the emission region size. GRB comes from the collimated relativistic outflow towards line of sight.

Multi-wavelength observations now reveal at least two distinct physical origins of cosmological GRBs. One type is initiated in the deaths of some special massive stars. The other is believed to be produced by the mergers of binary compact objects.

Although emission of γ-rays at even higher (TeV) energies has been theoretically predicted, it had not been previously detected.

Besides being strong emitters across the entire electromagnetic spectrum, GRBs are believed to be sources of non-electromagnetic signals, including cosmic rays, neutrinos, and gravitational waves. Recently the direct association between a GRB and a gravitational wave event has been made [2].

5.1.1 Prompt and Afterglow

The emission from GRBs occurs in two stages, which can partially overlap in time. The first stage is the burst itself or called 'prompt' emission phase, and the second is called 'afterglow' phase.

The prompt emission phase is conventionally defined as the temporal phase during which excessive sub-MeV emission is detected by the GRB triggering detectors above the instrumental background emission level. The duration is quantified by so-called the "T_{90}", which is defined as the time interval between the epochs when 5% and 95% of the total fluence is registered by the detector[1]. However, from theoretical point of view, the prompt emission and afterglow are better to be separated by its different origins, and it is widely accepted that the γ-rays are emitted at different physical locations. The most distinctive characteristic of the prompt emission is the light curve. Although it is a brief flash, the emission is intense and exhibits irregular variability on timescales shorter than milliseconds and lasts up to hundreds of seconds. The examples are shown in section G.1.

The ensuing 'afterglow' phase is characterised by emission that spans a broader wavelength range (from radio waves to GeV γ-rays) and decays gradually over much longer timescales compared to the prompt emission. The afterglow emission exhibits smooth power-law decay, and its spectrum can be explained mostly by synchrotron spectrum.

5.1.2 Short GRBs and long GRBs

T_{90} measured in the Burst And Transient Source Experiment (BATSE) energy band (25–350 keV) includes two Gaussian components with a separation line around 2 s [74] (for more details, see figure G.1). They are regarded as different classes and are called short GRBs and long GRBs.

The long GRBs have been clearly associated with the broad-line Type Ic Supernovae (SNe), which are the most massive cases of gravitational core collapse of a star together with the Doppler broadening of emission lines due to the larger velocity of the ejecta. Short GRBs are believed to be produced by the mergers of binary compact objects. The leading scenario is mergers of two neutron stars (NS-NS) or one neutron star and one stellar-size black hole (NS-BH).

Many evidences supports the associations, therefore the classification of Type I (compact-star-mergers) and Type II (massive-core-collapse) are introduced because the separation between the long and short populations is not clean. The associations are identified mainly by spectroscopy and localization as described below.

- Spectroscopy

  Some long GRBs have been detected with spectroscopically identified SN associations, all of which belong to Type Ic (e.g. [53, 103]). And even though the SNe are not spectroscopically identified, a light curve bump has been observed in the optical light curves of many long GRBs around 10 days after the GRB trigger, simultaneously changing its color. This

[1] Note that the duration varies dependent on the energy range of the detector. In this thesis I mention T_{90} measured in the Burst And Transient Source Experiment (BATSE) energy band (25–350 keV), if there is no description.

bump is consistent with the underlying thermal flux of supernova starting to overtake the synchrotron flux of GRB afterglow [31].

On the other hand, short GRBs do not show SN associated light. In the model of compact-star-merger, nucleosynthesis like r-process generates heavier elements and the photosphere would be at larger radius and lower temperature. Thus bright infrared emission rather than optical is expected. Such associated excess after around 10 days, called kilonova, r-process nova or mergernova, were detected for some short GRBs (e.g. [24, 55])

- Host galaxies and locations in host galaxies

  The studies show that the long GRBs occur in the environment with high star-forming rate, and it is quite consistent with the massive star origin of long GRBs. The majority of the long GRBs host galaxies are star-forming galaxies [49]. Long GRBs favor a low-metallicity environment [58]. Besides, most long GRBs reside in the brightest core regions of the host galaxies, where the specific star formation rate is the highest [49].

  On the other hand, the host galaxies of short GRBs are somewhat larger, the stellar population is relatively older, and metallicity is relatively higher [25]. As for the location in the host, most short GRBs are found to be far from the bright light of the host galaxies, and the physical offset of the GRB with respect to the center of the host is large [48]. And some are even hostless, suggesting that they may have been kicked away from host, or reside in distant faint host galaxies [23]. All these are consistent with the expectation of the compact-star-merger models.

- Gravitational wave

  The first NS-NS gravitational wave event GW170817 was associated with a short GRB, GRB170817A [1, 2]. The GRB was discovered also to be associated with kilonova events (e.g. [112])

5.1.3 Relativistic outflow

Because of the bright and high energy emission observed, together with fast variability, the emission region must move relativistically towards the observer [96]. Otherwise, calculating from total "isotropic" energy and the scale of emission region obtained from the burst duration, the photon density at the production site would be high enough to enable pair conversion, and the $>$ MeV photon would not be able to escape from the source. This requires the existence of a relativistic jet for the emission, and it is supported by the existence of the jet break in the afterglow light curve. The Lorentz factor of the bulk motion of jet is denoted as Γ.

The formation of the jet is still subject of intense debate [77, 117]. A widely used model is the matter-dominated fireball, which consists of baryons (primarily protons and neutrons), electron and positron pairs, and photons. A progenitor star undergoes a catastrophic event (a merger of binary or an explosion of massive star), resulting in a sudden release of gravitational energy. This creates a central engine and powers an outflow for a certain duration of time, during which gravitational energy or spin energy is released in the form

of thermal energy or Poynting flux energy, respectively. Later on, the energy is converted to kinetic energy of the outflow, making the ejecta reach a relativistic speed.

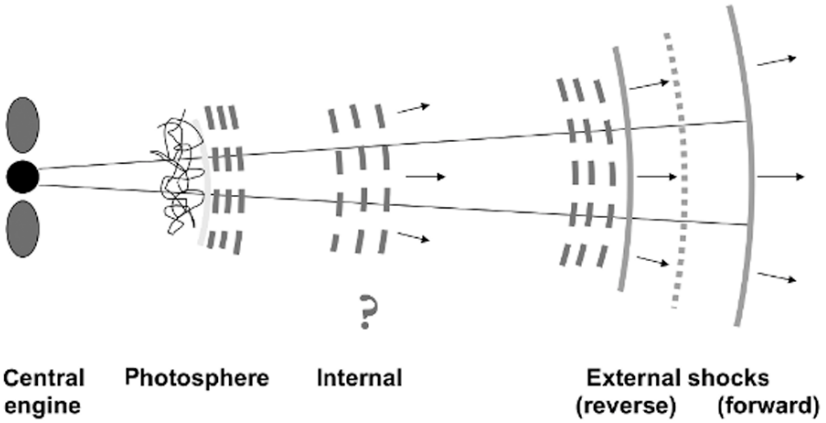


Figure 5.1: The evolution of a GRB jet within the general theoretical framework of GRBs

A cartoon picture of the evolution of a GRB jet within the general theoretical framework of GRBs. The dashed curves denote possible internal emission sites for the prompt emission, which are bracketed by the photosphere (included) and the external shocks (excluded). The afterglow is normally considered to be from forward shock in external shock region. Figure extracted from [117].

5.1.4 Prompt emission

Although the prompt γ-ray emission was the earliest detected signal more than five decades ago, the origin of the prompt emission is not well understood, and is still subject of intense debate.

The location of prompt emission is still not settled. However, it is indicated to be normally of internal origin (see figure 5.1). The innermost location is on the photoshere. It is where the photon optical depth for Thomson scattering becomes less than unity in the outflow.

A prompt emission spectrum may include three components: one thermal, and two non-thermal components (see section G.1). One of the non-thermal components is called "Band" component, and the other is called the high-energy component, because it extends to higher energies and it starts somewhat later (with respect to the thermal and the Band component). The significance of different spectral components may vary among GRBs. The physical origins of the three elemental spectral components have not been fully identified. A plausible picture attributes the thermal component to the photosphere emission, and the Band component to the non-thermal synchrotron radiation in the optically thin region. The high energy component remains unknown and subject to debate, even though some sort of Inverse Compton scattering processes are likely at play.

5.1.5 Afterglow emission

The long-lasting afterglow emission originates from shock waves caused by the interaction of the jet with the circumburst medium.

The shock waves are called 'external shocks'. There are two types of external shocks (figure 5.1); the forward shock propagates into the circumburst ambient medium, while early on a reverse shock propagates into the jet itself and crosses the jet in a short duration of time. The afterglow emission comes from these external shocks. In the beginning, the reverse shock coincides with the forward shock and later on the forward shock emission remains[2].

Spectrum

The ubiquitous property of a GRB afterglow is its "multi-wavelength" nature. As predicted from the synchrotron/SSC external shock model (e.g. [42, 85, 98, 99, 118]) , the afterglow should cover a very wide frequency range, from low-frequency radio to the TeV range. In the models, the broad-band afterglow spectrum emerges in the following way and is supposed to be a broken power law (details are given in section G.2.2).

In the relativistic shocks owing to plasma instabilities or other mechanisms, the magnetic field is efficiently amplified from values of a few μG, which are typical of the unshocked ambient medium. The electrons (and protons) would be accelerated, giving rise to bright broad-band non-thermal emission through synchrotron radiation. It can explain the afterglow emission of previously observed GRBs mostly in wide ranges from radio frequencies to GeV energies, and the spectral shape such as breaks originates from the properties of shocks. Since the energy carried by an electron is much smaller in reverse shock, the reverse shock emission actually peaks in IR/optical/UV, while the forward shock has the emission peaks in X-rays or even soft γ-rays [73].

However some theoretical scenarios predict emission at still higher energies towards TeV range, produced by additional radiation mechanisms. Among all them, one that is particularly important is IC radiation, whereby the population of electrons responsible for synchrotron radiation also give rise to photons at higher energies by Compton upscattering ambient low-energy photons. This is called SSC. IC emission has been clearly observed in other types of cosmic sources of electron synchrotron emission, such as pulsars or blazars.

Until very recent, the highest energy at which a GRB had been previously seen was ~ 95 GeV, with a single photon detected from GRB 130427 by *Fermi*-LAT [9]. This is close to the maximum photon energy that is theoretically possible due to the synchrotron process in the afterglow, called synchrotron burnoff limit (section G.2.2). However, its quantitative significance has been uncertain because the SSC luminosity and spectral properties depend strongly on the poorly constrained physical conditions in the emission region, like the magnetic field strength.

Given the spectral shape of both the synchrotron and SSC components, the physical properties in the shock region can be revealed. Moreover, the time-resolved spectra can further lead us to the time development of the shock that decelerates in the ambient medium.

[2] If the central engine is long lived, or if Γ in the ejecta is stratified (widely distributed), the reverse shock can be long lived.

Light curve

The evolution of afterglow is typified by a power-law smooth decay in time in all wavebands, and it is well explained by theory, in which the jet is decelerated by a circumburst ambient medium, as the relativistic forward shock propagates into the medium. The theory of relativistic blast waves has been worked out in a classical paper by Blandford & McKee in 1976 [30], and later on the afterglow models were proposed [72, 85, 92] before the actual observation of the afterglow.

However further multi-wavelength observations revealed a more complex chromatic behavior, hence suggesting more complicated afterglow physics. The afterglow is a superposition of multiple emission components. For example,

- The afterglow emission is overlapped by the internal origin prompt emission in the early phase of afterglow.
- Early on, a reverse shock occurs and the emission from reverse shock sometimes coincides with the emission from external shock, during reverse shock crossing phase.
- A steepening temporal break attributed to the end of collimation of GRB jets occurs. It is called "jet break".
- In the optical afterglow light curves of a good fraction of long GRBs, a bump feature, usually with a red color, shows up about a week after the GRB trigger. This is usually interpreted as the signature of an associated supernova, with thermal emission dominating the synchrotron emission.

Among the wide range of the spectrum, the most important waveband in relation with the TeV energy range is X-ray, because the TeV energy is expected to be from SSC, which is based on the synchrotron emission peaked at around X-ray. The morphology of the X-ray afterglow light curve also indicates multiple components, whose existence and characteristics vary from burst to burst. Many are with just a single power-law decay, with the index of -1, which is called "normal decay". It may contain steeper decay because of the jet break in the end. Some show very steep decay as the tail of prompt emission. Some show a shallower decay in the beginning or after the steep decay, which could be explained by the energy injection from a spinning-down newly-born magnetar or from accretion caused by core-envelope structure [26]. They are summarized in section G.2.1.

5.1.6 GRBs and MAGIC

As mentioned above, TeV γ-rays from GRBs have been predicted. If the IC component (in addition to the synchrotron component) is measured, it allows a better understanding of the physical parameters of the source such as the magnetic field strength and particle density, which is limited when observing only the synchrotron emission.

Therefore the search for TeV γ-rays from GRBs has been one of the more fascinating challenges in ground-based γ-ray astronomy during the last 30 years, however their detection had remained elusive. As GRBs occur at unpredictable times and sky locations and then rapidly fade in brightness, observing them requires telescopes with either a very wide field of view, or the capability to

quickly respond to alerts from satellite instruments with wide fields of view (and optimized for the search of GRBs).

Since their inception, MAGIC has been regularly conducting automated follow-up observations of GRBs in response to alerts from space-borne instruments. In addition to the fast slewing speed of the telescopes, to achieve the fastest possible response to GRB alerts, an automatic alert system (AAS) has been also developed. If an alert is tagged as observable by the AAS, the telescopes automatically repoint to the new sky position and the new observation starts. After repointing the telescopes to the different direction, the condition of sky and the telescope structure are different, thus the relevant procedures[3] are also automated. To maximize the chance to detect GRBs, there has been many years of technical improvements and dedicated efforts for various points. One of the most important attempts to increase the chances to observe and detect GRBs was to adapt the observation and analysis to relatively high moon light condition. Since GRBs occur at random times and locations in the sky, increasing the number of possible observations was crucial.

At last, the MAGIC telescopes were able detect TeV emission from GRBs. In 2016 there was already marginal detection at 3 standard deviations of GRB 160821B by the MAGIC collaboration. And finally in 2019, the first clear detection was announced with GRB 190114C at more than 50 standard deviations.

.

5.2 Observation of GRB190114C

5.2.1 MWL observation history on GRB190114C

GRB 190114C was first identified as a long-duration GRB by the Burst Alert Telescope (BAT) onboard the Neil Gehrels Swift Observatory (*Swift*) [60] and the Gamma-ray Burst Monitor (GBM) instrument onboard the *Fermi* satellite [62] on 14 January 2019, 20:57:03 universal time (hereafter T_0)[4]. The time t indicates the elapsed time from T_0.

At $t = 22$ s, *Swift*-BAT distributed an alert reporting the first estimated coordinates of GRB 190114C. The AAS of MAGIC telescopes received the alert, validated it as observable, and triggered the automatic repointing procedure. The telescopes repositioned the pointing position by slewing in fast mode and reached the target position after approximately 27 s. The telescopes were on target, and began tracking GRB 190114C at $t = 50$ s. Data acquisition started at $t = 57$ s and the data acquisition system was operating stably from $t = 62$ s. The observation data with GRB 190114C was taken from $t \sim 57$ - 15,912 s. The automated observation was executed although the GRB happened during the presence of moon, by which the Night Sky Background (NSB) was 6 to 8 times higher. The regular operation under presence of relatively bright moon is possible only by MAGIC among the leading IACTs [11], which proved to be a crucial advantage for the GRB observation. Together with the relatively high

[3] The trigger tables are reloaded, appropriate electronics thresholds are reset, and the mirror segments are suitably adjusted by the Automatic Mirror Control hardware.

[4] Subsequently, it was also detected by several other space-based instruments, including *Fermi*-LAT, INTEGRAL/SPI-ACS, AGILE/MCAL, Insight/HXMT and Konus-Wind

Zd, the energy threshold of the observation was above 0.2 TeV, which is about four or five times higher than the energy threshold that MAGIC telescopes have for soft-spectrum sources observed in low zenith angles (below 35 deg) during dark nights.

In the first 20 minutes of the data, the online analysis showed the significance of the total γ-ray signal is more than 50 standard deviations. The significant detection in TeV opens a new window in the electromagnetic spectrum for the study of GRBs. Its announcement [86] triggered an extensive campaign of follow-up observations.

In the end, the follow-up observations gathered for the publication [7] were from instruments onboard six satellites and 15 ground telescopes (radio, submillimetre, near-infrared (NIR), optical, ultraviolet (UV), and VHE γ-rays) for the first ten days after the burst. The frequency range covered by these observations spans more than 17 orders of magnitude, from 1 to about 2 $\times 10^{17}$ GHz (from 5 $\times 10^{-6}$ to about 10^{12} eV), the most extensive for any GRB to date.

5.2.2 Basic properties obtained from MWL observations

The basic properties of GRB 190114C obtained from multi waveband observation are as follows [7, 8].

- T_{90} and prompt emission
  Its duration in terms of T_{90} was measured to be about 116 s by GBM onboard *Fermi* [62] and about 362 s by *Swift*-BAT [76]. Judging from the moment the last flaring-emission episode ends, the prompt emission lasted for approximately 25 s.

- red-shift
  z = 0.4245 $\pm$ 0.0005, corresponding to cosmic distance [39, 101].

- Isotropic-equivalent energy of prompt emission
  The isotropic-equivalent energy of the emission at energy of $\varepsilon = 1$–10^4 keV during T_{90} observed by *Fermi*-GBM was $E_{\gamma,iso} \sim 2.5 \times 10^{53}$ erg, implying that GRB 190114C was fairly energetic, but not exceptionally so compared to previous events.

- Afterglow onset time and reverse shock duration
  The onset of the afterglow component is estimated to occur from around $t \sim$ 5-10 s, because there is a term of quiescence at $t \sim$ 5-15 s during which the flux decays smoothly and similarly as $t >$ 25 s (seen in *Swift* BAT and *Fermi* GBM). The decay of flux F follows a power law as a function of time, $F \propto t^{-\alpha}$ where α is energy range dependent but comparable to typical values as the normal decay phase shown in figure G.5. The reverse shock component can be seen up to $t \sim 1000$ s for optical and up to $t \sim 10^5$ s for radio emission.

- Initial bulk Lorentz factor
  From the afterglow onset time and the ambient medium density, the initial bulk Lorentz factor are constrained. Typically there are two kinds of assumptions for the ambient medium density, which are $s = 0$ (homogeneous medium) and $s = 2$ (wind-like medium, typical of an environment

shaped by the stellar wind of the progenitor), in the radial profile defined as eq.(G.2.2),

$$n(R) = n_0 R^{-s}$$
$$n_0 = 3 \times 10^{35} A_*$$

The detail is described in section G.2.2.

The initial bulk Lorentz factor for those two cases are $\Gamma_0 \sim 300$ and $\Gamma_0 \sim 700$ for s = 2 and s = 0.

- Synchrotron peak shift
  The bulk Lorentz factor Γ decreases due to the deceleration of jet. Correspondingly the synchrotron peak energy shifts towards low energy and the light curve shows decay-slope change with time; if the peak is at higher energy than the observation energy range, the light curve slope steepens and the spectrum within the detection range changes from hard through flat to soft when the peak moves across the range. This feature can be seen in the NIR-optical light curves at $t \sim 10^5$ s, which is relatively large, supporting the existence of the peak around 10 keV in the beginning.

- Fast and slow cooling
  In a wind-like scenario, X-ray and optical emission and their evolution in time can be explained if $p = 2.4$–2.5 and the emission, which is initially in the fast-cooling regime, transitions to a slow-cooling regime around $t \sim 3 \times 10^3$ s (See section G.2.2).

5.3 The analysis of MAGIC data

The MAGIC data were analysed in detail using MAGIC Analysis and Reconstruction Software (MARS), the standard MAGIC software. The analysis was carried out by the team of people appointed to the GRB190114C observation (from now on, I will refer to them as GRB analysis team). The GRB analysis team officially adopted my novel energy estimation method, because it showed superior and more reliable performance than LUTs-Erec, the previous standard method for energy reconstruction. However I show the alternative analysis done by myself for two reasons. Firstly, the GRB analysis team used the older version (RF-Erec-v3) of the energy estimation, because it was the latest version I had released at the moment. In this thesis I discuss the result using the final version (RF-Erec-v5). Secondly the event selection applied for the spectral analysis was very conservative, and suppressed the statistical abundance, causing the loss of the highest energy data points. Thus the benefit from the novel method is not maximized. Moreover, I apply an alternative event selection to the GRB data set, yielding better results.

In the following, section 5.3.1 describes the simulation data and image cleaning to match the observation condition, which are common to both my alternative analysis and the original analysis by the GRB analysis team. Section 5.3.2 describes the difference of the event reconstruction and event selection (cut) between the two analyses. This chapter aims at showing the effect of RF-Erec-v5 energy reconstruction in the improvement of spectral analysis compared to the LUTs-Erec. The comparison between the my alternative analyses with

RF-Erec-v5 and the analysis with RF-Erec-v3 used by GRB analysis team is shown in Appendix I.

5.3.1 Observation conditions and analysis strategy

In the observation, the MAGIC telescopes tracked the source in the wobble mode by 0.4° offset. There are two issues to investigate carefully in the analysis.

First, the zenith distance(Zd) of the observation was relatively high, about 55° at T_0. At the Zd larger than 50°, the Zd dependence of the shower images is high and also the azimuth dependence becomes relevant, therefore a set of generic simulation data was produced for better accuracy of the analysis with the same tracking positions[5], although there is standard simulation data available. The GRB location was almost at its culmination in the sky at t=0, and the Zd changed only within 1° in the first 1 hour ($55.4° < \mathrm{Zd} < 56.4°$). Therefore the energy threshold should be stable with Zd. The other conditions relevant to the energy threshold, like night sky background level, discriminator threshold and trigger rage were also stable.

Second, the GRB was observed under the presence of the moon. In this case, the sky is brighter and the contamination of the NSB photon is higher. Therefore the discriminator threshold (DT) during the observation was automatically set higher to suppress accidental trigger. Moreover, the images are contaminated with the NSB photons, and the image cleaning needs a higher cleaning level. To deal with these issues, the analysis followed the steps optimized for the data taken under moderate moon illumination [11]. As mentioned in section 3.6, there are two threshold values for the image cleaning. The higher cleaning threshold value compared to standard analysis, $(Q_1, Q_2) = (9, 6.5)$[phe], was applied. Correspondingly, the simulation data also needs to be processed in the same manner. Before the same cleaning level to be applied, the simulation data needs be produced with the same NSB noise level and the DT level. The GRB analysis team produced the simulation data with different DT settings (DT =4,5,7 phe) and corresponding noise levels, and they chose the one closest (DT=7) to the real data.

contains the hadronic background events. To remove them,

The consistency between the observation data and the simulated data was checked over the distributions of Size for γ-ray-like events, because it is the most strongly correlated variable with the γ-ray energy. Figure 5.2 show the comparison, and they are consistent over 70 phe. The simulation data may be systematically lowered by about 10 phe. However the GRB analysis team showed the spectra using these simulation data with different DT settings are consistent each other. Therefore the simulation data with DT=7 phe would be good enough. Moreover, if Size $>$ 65 is applied, the distributions become consistent.

For the Size cut of the analysis, the peak of the distribution about 80 phe would be proper. However, based on the aforementioned reason, the Size cut could be lowered down to 65. The events with such low Size are dominated by low energy events, therefore they contribute to lower the energy threshold.

[5] In the MC simulation, the observation direction is uniformly distributed in the generation of the data, where the Zenith distance angle (Zd) is from 55 - 62 deg and Azimuthal angle (Az) is 175 - 212 deg, to be close to the observation of GRB190114C.

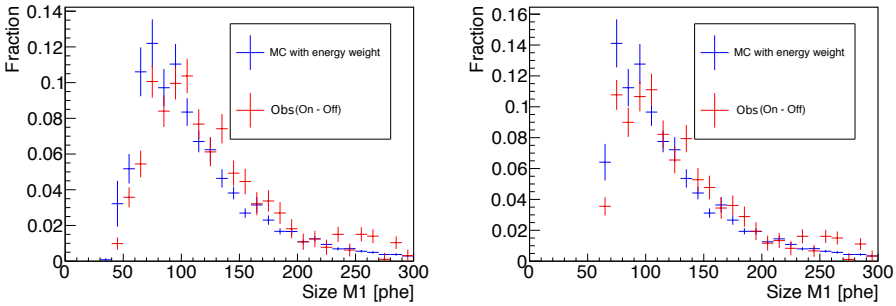


Figure 5.2: Normalized distribution of the parameter Size for simulated and real γ-ray events, before and after Size cut

The two panels show the normalized distributions of Size, before (left) and after (right) the Size cut is introduced. The Size cut removes the systematic difference by DT settings, between γ-ray events by MC simulation (blue) and the γ-ray candidate events from the observation data (red). The simulated events are drawn with energy-spectrum weight applied to the true energy. The energy spectrum could be determined via tentative analysis and was power law with the index -5, which was later on confirmed as consistent with the detailed analysis (see table 5.2). Observation data are drawn as the subtraction of the histograms as the "ON" histogram (the events whose reconstructed incoming direction is within $\theta^2 < 0.1$ deg^2 around the source position) minus the "OFF" histogram (the ones around counterpart position of the wobble). The hadronnes cut (Hadronness < 0.5) is also applied to both the simulation and the observation data.

5.3.2 Event reconstruction and event selection

To both analyses, the same procedure is applied before event reconstruction, including the production of simulation data. Here I describe the differences between my alternative analysis and the one performed by the GRB analysis team, which relate to the event reconstruction and event selection (cut) between the two analyses.

GRB analysis team

The energy reconstruction was performed through the novel method with RF. This is the first time the novel method was applied in a MAGIC publication. However, the version I could provide at that moment was version 3 (3.1 for LZd and 3.2 for HZd), thus ver 3.2 was applied (hereafter RF-Erec-v3). The difference of the versions of the energy reconstruction is described in section 4.4.7.

The events are selected under the cut condition on Size > 80, hadronness efficiency 0.9, θ^2 efficiency 0.75 and EnergyRMS< 0.3. For the discussion on these values, see also section 3.10.3. About the role of EnergyRMS, see section E.1).

Alternative analysis

I adopt the RF-Erec-v5 energy reconstruction. I also optimize the cut condition to be suitable for strong signal with very soft spectrum. The flux is very abundant and its spectrum is far softer than that of usual sources. Therefore careful consideration needs to be given to the suitability to the data. I adopt the cut condition as Size > 65, hadronness efficiency 0.9, θ^2 efficiency 0.85. I do not apply any cut on the variable EnergyRMS. See section E.1 for details.

The optimised condition intends to deal with extraordinary high flux with very steep energy spectrum, which had not been experienced in MAGIC to date. The lower Size cut tries to mitigate the systematic effects in the lowest energy, and the higher θ^2 efficiency is to gain more statistics at the highest energies. The detailed discussion and the comparison with the analysis by the GRB analysis team is summarised in appendix I.

For the spectral analysis with the LUTs-Erec, the same cut condition as for RF-Erec-v5 is applied.

5.3.3 Intrinsic and observed spectrum

Intrinsic spectrum

The aim of the spectral analysis of the observation data is to obtain a reasonable estimation of the γ-ray spectrum when it was emitted, referred to as the 'intrinsic spectrum'. However γ-rays from cosmologically distant objects such as GRBs suffer from the attenuation by the EBL. The attenuation is energy dependent and quickly rises for the energy higher than a few hundred GeV. This effect is also dependent on the distance. If the γ-ray source is at cosmological distance, the observed γ-ray spectra can be substantially modified. Therefore the observed spectrum needs to take into account the EBL effect, if the interest is in the intrinsic spectrum. There are some models of the attenuation available, and in this study the model of Domínguez et al. [44] is used[6]. The detail about EBL absorption is described in section G.3.

The reasonable application of unfolding

Since the attenuation is strongly energy dependent, the forward and unfolding are not equivalent to obtain the intrinsic spectrum. This is specially important in the GRB data, because the source is at cosmological distance and the energy dependence is so strong that the observed spectrum becomes very steep. Moreover, the flux is so high that the observed spectrum extends above 1 TeV, which means the spectral analysis needs to deal with the attenuation factor changing from about 1 to 300. Therefore this unusual situation requires a careful treatment in determining the intrinsic spectrum through unfolding.

The suitable method is forward folding. It can directly show the consistency of the assumption of intrinsic spectrum with the observed number of excess events, because the observed spectrum can be analytically calculated from an intrinsic spectrum when the attenuation is given as a function of energy. On the other hand, the unfolding is to obtain only the observed spectrum. The calculation in binned values from observed spectrum to intrinsic spectrum suffers from additional uncertainty.

Therefore, I perform the spectral fitting of intrinsic spectrum only by the forward folding. The unfolding is used only for obtaining the estimated flux in the true energy bins, to check the consistency with the forward folding results.

[6] In the spectral analysis, the uncertainty on the EBL model, as well as the choice of the EBL model, contributes to the systematic error on the determination of the parameters of the intrinsic spectrum model. These contributions are investigated. The systematic error from the choice of EBL model is less than the systematic uncertainty in a model. The systematic uncertainty in a model is smaller than the statistical uncertainty (one standard deviation). The details about the investigation is described in [8].

5.4 The improved spectral analysis

In this section, I compare the spectrum obtained from the latest version of the RF energy reconstruction (RF-Erec-v5, hereafter RF-Erec in this section), with the classical energy reconstruction in the MAGIC collaboration (LUTs-Erec). This is the method that would have been used if my energy estimation method had not been available.

In forward folding, the maximum likelihood fit process also takes into account the EBL absorption by setting intrinsic spectrum as the model spectrum together with the redshift $z = 0.4245$, as mentioned in section 5.3.3. The intrinsic spectrum to fit is chosen to be simple power law,

$$\frac{dF}{dE} = f_{\mathrm{EBL}} \left[A_0 \left(\frac{E}{E_0} \right)^{\Gamma_{ph}} \right] \sim A_{0,\mathrm{obs}} \left(\frac{E}{E_0} \right)^{\Gamma_{ph,\mathrm{obs}}}, \tag{5.4.1}$$

where dF/dE [$\mathrm{TeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$] is the differential flux as a function of energy E, measured after the intrinsic spectrum is distorted by the EBL absorption f_{EBL}. As explained in section G.2.2, a power law is a reasonable functional shape for the intrinsic spectrum. A_0 is the amplitude, Γ_{ph} is spectral index, and $E_0 = 500$ GeV is the pivot energy[7]. As will be shown later, the actual measured spectral shape distorted by EBL absorption can still be approximated as a power law. $A_{0,\mathrm{obs}}$ and $\Gamma_{ph,\mathrm{obs}}$ are the power-law amplitude and spectral index for the measured power-law spectrum.

In unfolding, the bin range for estimated energy and true energy needs to be selected. I adopted the same criteria as in section 4.6.4, in the setup to analyze the power-law spectrum with index -5. The selected ranges are shown in figure 5.3. In addition, unfolding needs to control the regularization strength. The criteria for the best strength are chosen to balance the agreement with the measured excess events and the noise. More details about the strategies to control the regularization strength are described in appendix H.

5.4.1 The result of forward folding

The spectral parameters of the intrinsic spectrum obtained from the forward folding are shown in table 5.1. The spectrum obtained with LUTs-Erec is significantly steeper than that from RF-Erec. And the uncertainty of the slope is larger in LUTs-Erec. The numbers of degree of freedom (NDF) are also different.

Table 5.1: Forward folding results for the GRB190114C data

	A_0 [10^{-9}/TeV cm^2 s]	Γ_{ph}	χ^2/NDF
LUTs-Erec	$5.99^{+0.92}_{-0.93}$	$-2.91^{+0.28}_{-0.33}$	6.8/8
RF-Erec	7.60 ± 0.67	$-2.32^{+0.17}_{-0.19}$	14.1/6

[7] If E_0 is adjusted to de-correlate A_0 and Γ_{ph}, E_0 becomes different between RF-Erec and LUTs-Erec and A_0 cannot be compared. Therefore E_0 is fixed to the value of 0.5 TeV, which is close to the de-correlation energy for both methods. The de-correlation energies are 444 GeV for RF-Erec and 414 GeV for LUTs-Erec.

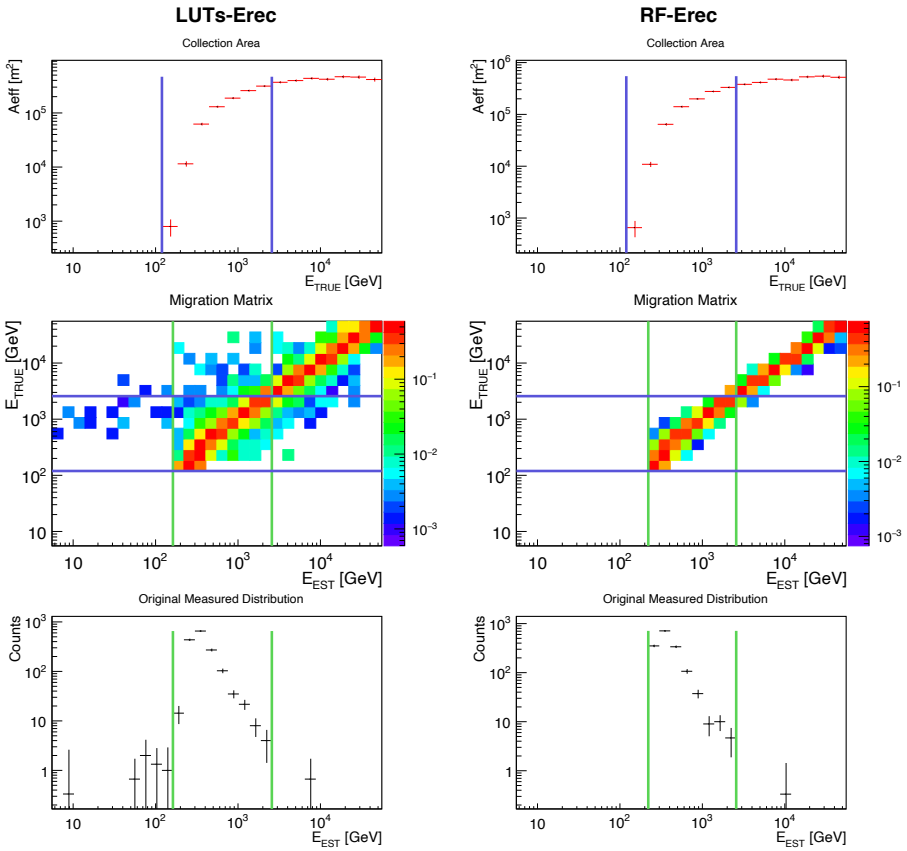


Figure 5.3: Collection area, migration matrix and excess events of GRB190114C with LUTs-Erec and RF-Erec

The ranges for calculating the unfolding are also shown as the colored bars in the plots. The green (blue) bars are the lowest and highest edge of the estimated (true) energy range. In the unfolding with LUTs-Erec, both Bertero methods were forced to reduce the selection of the estimated energy bins by one in the highest energy side, because they did not converge regardless of the criterion of choosing the best regularization strength.

More investigations clarify the situation. From the top 4 rows in the control plots presented in figure 5.4, both RF-Erec and LUTs-Erec show consistency between the fit result and the observed counts. The upper two rows report the number of ON events and OFF events in which the observed and expected (by the fit) counts are drawn together. The next two rows show the difference between the observed and expected, and the χ^2 contribution from all the energy bins.

However there is clear difference in the contamination fraction shown in the bottom row of the control plots, figure 5.4. The energy bins for LUTs-Erec shows very high contamination fraction, quickly rising with energy from 0.5 at a few 100 GeV and reaching already almost 1 at a few TeV. This is resulted from the noisy migration matrix, shown in the second row of figure 5.3, together with the very steep observed spectrum. Even though it is the tiny fraction of low energy events to spill into high reconstructed energy, it becomes dominant (high contamination fraction) in the high energy bins.

Consequently, a clear difference appears in the reconstruction of the flux data points, shown in figure 5.5. In the forward folding with LUTs-Erec, the high contamination fraction indicates that the excess events in an energy bin are dominated by such spillover events from lower energies. The flux data points are dragged towards lower energy due to too low estimated median energy, and many of them cannot be placed. This significant reduction in the number of data points is in the similar way to the performance evaluation discussed in section 4.6.4. On the other hand, the forward folding with RF-Erec provides reliable data points until a few TeV without being affected by the contamination.

The noisy migration matrix with LUTs-Erec can also explain the difference in the fitted parameters of the intrinsic spectrum. Since the outliers (low-probability cells) of the migration matrix come from the minority of the population, they cannot be reproduced by the limited statistics of the test samples of simulated γ-rays, even if the simulated γ-rays describe perfectly well the real γ-rays. Consequently the fluctuated pattern of outliers in the migration matrix would not lead to the original energy distribution. Instead it is likely to find a consistent answer to migrate the high estimated-energy events back to low energy events, because of the noisy distribution. Therefore it is natural for a LUTs-Erec to obtain steeper intrinsic spectrum, by losing high energy events 'absorbed' into low energy. This is in apparent contradiction to the performance evaluation discussed in section 4.6.4, where the spectral slope difference is not seen. However, this result is expected because I had to use the same set of events for generating the migration matrix for analysis, as well as that for generating an simulated observation data.

The noisy migration matrix with LUTs-Erec also means that there are more possibilities to find a consistent migration, thus the uncertainty is naturally higher. Moreover, the noisy migration matrix also explains the larger NDF in the fit, which means that the excess events distribute in more bins of estimated energy. The larger NDF in this sense is not welcomed, because it comes from smearing the distribution, which means loss of information.

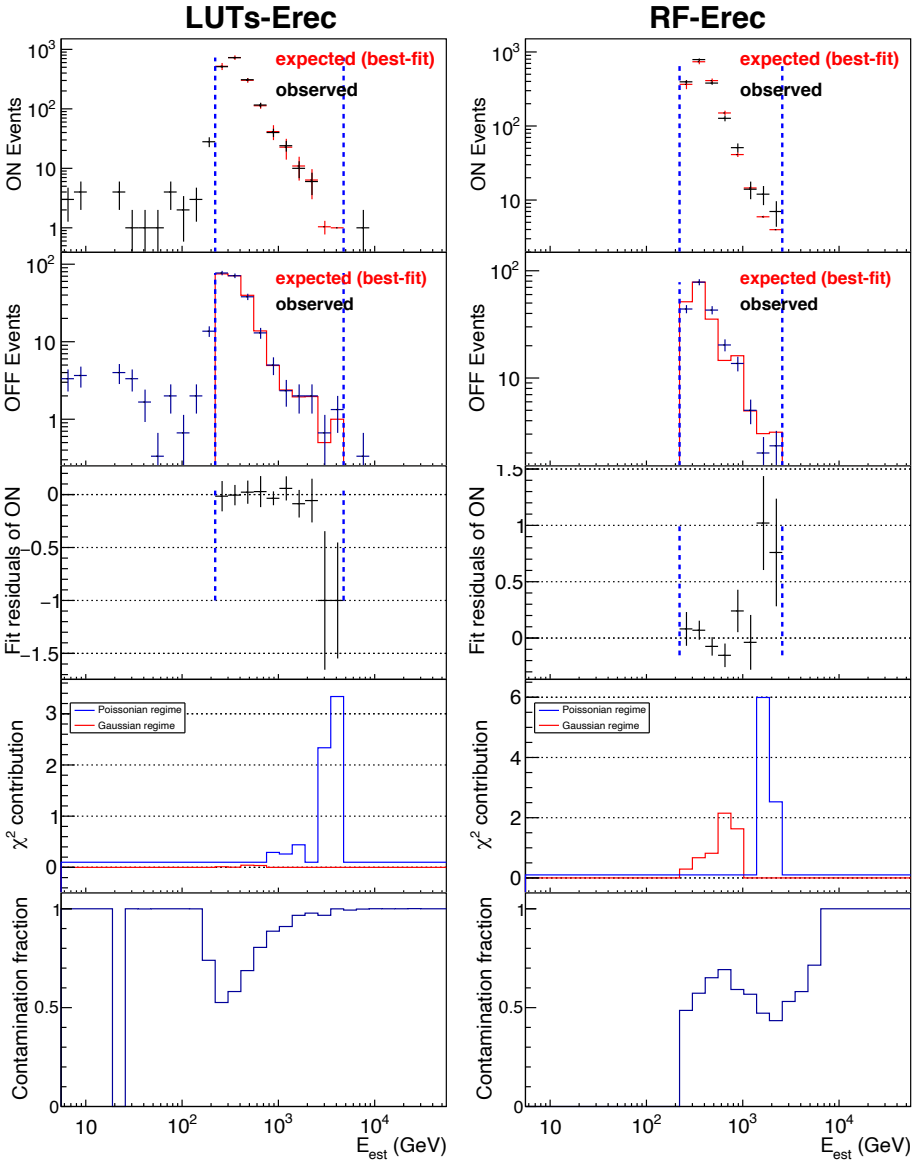


Figure 5.4: Control plots from forward folding with LUTs-Erec and RF-Erec on GRB 190114C data

Control plots yielded by the forward folding when using the RF-Erec and LUTs-Erec energy reconstruction on the VHE MAGIC data from GRB190114C. See the caption of figure 4.42 for a detailed description.

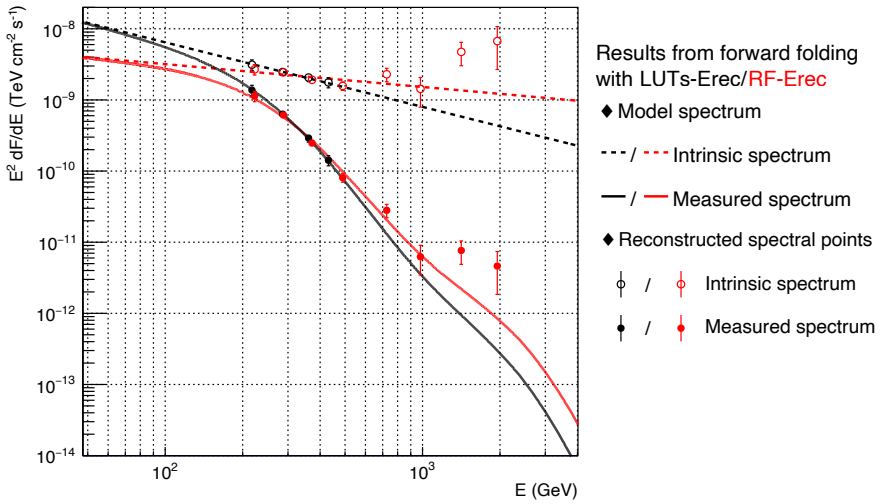


Figure 5.5: VHE spectrum of GRB19014C reconstructed by forward folding, with LUTs-Erec and RF-Erec

The reconstructed spectra by LUTs-Erec and RF-Erec are drawn overlaid in different colors, shown in the legend. The lines are the fitted model spectra. The dotted line is the intrinsic (deabsorbed) spectrum as a power-law function, the solid curve is the observed (EBL-absorbed) spectrum. The points are the reconstructed flux, using the spectra for taking into account the energy migration in the collection area. The filled circles are the reconstructed flux data points as the measured spectrum, and the open circles are the flux level after EBL correction. The parameters of the spectral model obtained by forward folding and their goodness of fit are shown in table 5.1.

5.4.2 The result of unfolding

Evaluation of the results

The results by unfolding are shown in figure 5.6. The unfolded spectra are the most reasonable results among different criteria for choosing the best regularization strength, described in section H.1, where both χ^2_0 and the noise are close to reasonable values.

The unfolded spectra with RF-Erec are consistent among the different regularization methods and also with the observed spectral curve from forward folding. However the ones with LUTs-Erec have different tendencies among different regularization methods. The highest energy data point is isolated from the lower data points with a gap in Bertero(W), and the flux in the lowest energy data point is low in Schmelling.

The χ^2_0 in unfolding, shown in the right of table 5.2, tells how well the unfolded spectral points agree with the measured number of excess events through the migration matrix. The values are low in both RF-Erec and LUTs-Erec, thus all the unfolding give the results reasonably explaining the measured counts.

Table 5.2 also shows the result when they are fitted with a power-law function. The spectral points obtained by unfolding can be regarded to follow power law, based on the forward folding results, shown in the bottom row of the table.

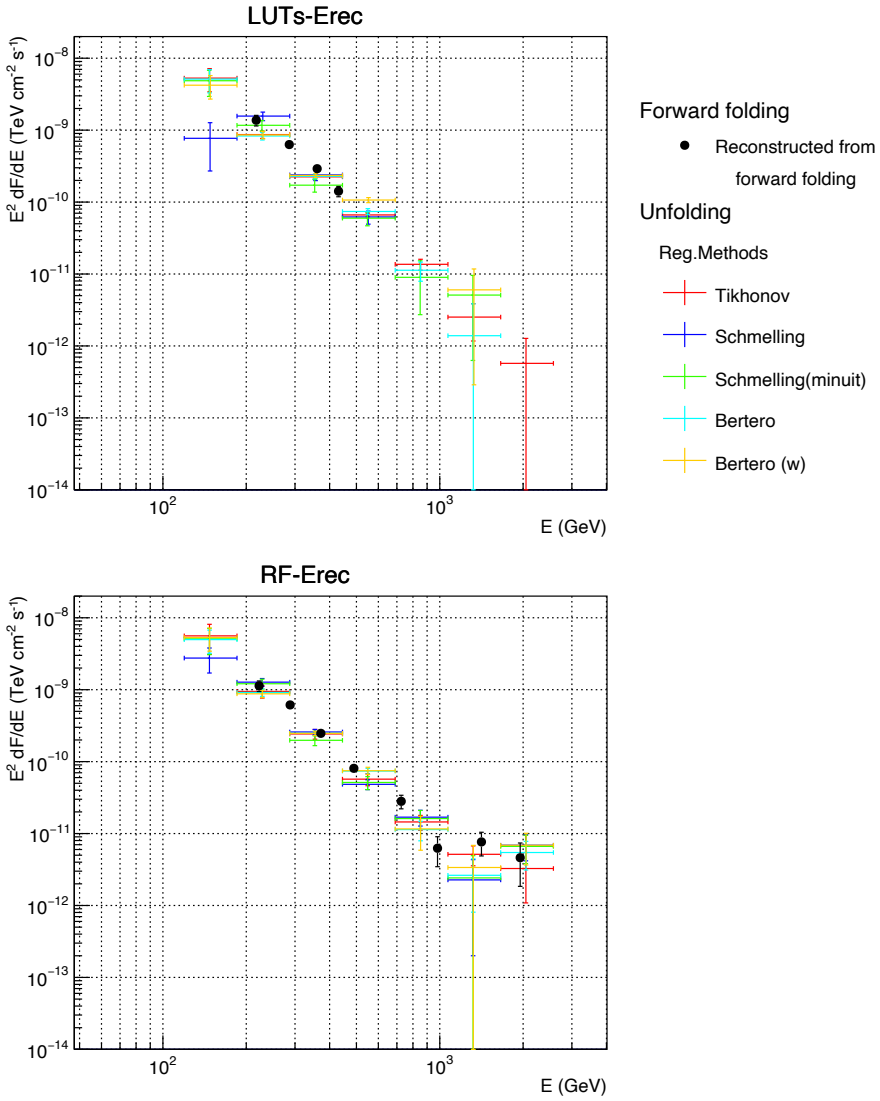


Figure 5.6: VHE spectrum of GRB19014C reconstructed by unfolding, with LUTs-Erec and RF-Erec

The results are separately drawn for the ones with the LUTs-Erec and RF-Erec. The colored points are the reconstructed flux by unfolding with different regularization methods, shown in the legend. The black points are the reconstructed flux based on forward folding.

Table 5.2: Results from unfolding and power-law fits to unfolded data points of GRB190114C data with LUTs-Erec and RF-Erec

(a) LUTs-Erec

	Unfolding	**Power-law fit**		
	χ^2_0/n_y	$A_{0,\mathrm{obs}}$	$\Gamma_{ph,\mathrm{obs}}$	χ^2_{PL}/NDF
Tikhonov	7.6 / 9	2.99 ± 0.39	-5.30 ± 0.26	0.6 / 5
Schmelling	8.5 / 9	4.07 ± 0.32	-4.58 ± 0.15	23.5 / 2
Schmelling(m)	9.3 / 9	2.62 ± 0.33	-5.48 ± 0.25	2.7 / 4
Bertero	6.8 / 8	3.26 ± 0.24	-5.05 ± 0.15	8.6 / 4
Bertero(w)	9.9 / 8	4.69 ± 0.35	-4.34 ± 0.17	11.3 / 3
Forward folding		2.98 ± 0.40	-5.58 ± 0.28	8.7 / 8

(b) RF-Erec

	Unfolding	**Power-law fit**		
	χ^2_0/n_y	$A_{0,\mathrm{obs}}$	$\Gamma_{ph,\mathrm{obs}}$	χ^2_{PL}/NDF
Tikhonov	6.6 / 8	3.27 ± 0.25	-5.19 ± 0.17	1.8 / 5
Schmelling	11.3 / 8	3.10 ± 0.18	-5.32 ± 0.13	10.6 / 5
Schmelling(m)	6.6 / 8	2.96 ± 0.24	-5.32 ± 0.19	7.7 / 5
Bertero	7.5 / 8	3.44 ± 0.22	-5.11 ± 0.13	9.0 / 5
Bertero(w)	7.7 / 8	3.48 ± 0.22	-5.04 ± 0.13	6.9 / 5
Forward folding		3.44 ± 0.25	-5.21 ± 0.16	10.0 / 6

The left column shows the consistency of the unfolding result, χ^2_0/n_y, where χ^2_0 is the chi-square between back-folded distribution of number of excess events and the one measured, and n_y is the number of bins of estimated energy which were taken into account in unfolding process (the reason for Bertero to have a lower n_y with LUTs-Erec is mentioned in the description of figure 5.3).
The right columns are the result of the power-law fit to the unfolded spectra, performed with the correspondent regularization methods, using RF-Erec and LUTs-Erec. In the bottom of each table, the fit result to the reconstructed points by forward folding after EBL absorption is also shown. The power-law function used for the fit is $\frac{dF}{dE} = A_{0,\mathrm{obs}} \left(\frac{E}{E_0}\right)^{\Gamma_{ph,\mathrm{obs}}}$, where $\frac{dF}{dE}$ [$\mathrm{TeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$] is the differential flux as a function of energy E, $A_{0,\mathrm{obs}}$ is the amplitude, shown in the unit of [$10^{-10}/\mathrm{TeV}\,\mathrm{cm}^2\,\mathrm{s}$] in the table, $\Gamma_{ph,\mathrm{obs}}$ is spectral index, and E_0 is the pivot energy set to be 0.5 TeV.

The reconstructed flux points of measured spectrum based on the forward folding results are fitted with a power law, and show the good agreement. Since the forward folding already shows good agreement of the model with the measured counts as shown in table 5.1, this result can also be validated.

Comparing the result of the power-law fit, the ones with RF-Erec show the consistent values of amplitudes $A_{0,\mathrm{obs}}$ and index $\Gamma_{ph,\mathrm{obs}}$ among all the regularization methods, and the goodness-of-fit are acceptable for all the results by the different regularization methods. However, LUTs-Erec gives significantly different results among each other, with some goodness-of-fit values that are not acceptable.

Evaluation of the process and the control plots

There are also the differences in the unfolding process itself together with the control plots. For the reliable result, the unfolding should show robustness in the choice of the best criterion of searching for the best regularization strength, denoted in section H.3, and the control plots should be with reasonable appearance. The detailed discussion about the differences is given in section H.4, and I summarise the discussion in the followings.

- If the same criterion of choosing the best regularization strength is selected, the unfolding with RF-Erec always gives better χ^2_0 than with LUTs-Erec for all the regularization methods.
- The unfolded spectra with RF-Erec show better agreement with each other, not only among the different regularization methods, but also among the different criteria of choosing the best regularization strength. On the other hand, LUTs-Erec showed inconsistencies among the spectra by different criteria of choosing the best regularization strength in all the regularization methods except for both Bertero methods.
- While RF-Erec went all fine in the convergence, LUTs-Erec had a problem in unfolding with Schmelling method, where the χ^2 minimization did not converge partially for the regularization strengths ω larger than $\sim 10^3$, which can be seen in the right panels of figure H.2d. In both Bertero methods the unfolding did not converge regardless of the criterion. This problem was solved by changing the selected ranges of energy bins, dropping the highest estimated energy bin.

It is notable that the performance difference of unfolding on the GRB data with RF-Erec and with LUTs-Erec is similar to that found on the simulated observation data for a steep spectrum, which I discussed in section 4.6.4; the LUTs-Erec showed worse χ^2_0, inconsistent results between different regularization strength choice and different regularization methods, failure of convergence, and deviation from reality. RF-Erec is superior than LUTs-Erec in all these areas.

5.5 The interpretation of TeV component

Here I summarise the results related to the TeV emission detected by MAGIC, which have been recently published in Nature [7, 8]. The temporal evolution of the flux in the MAGIC energy range suggests its connection to the afterglow, observed in the other wavebands. Therefore the emission in TeV energies should be explained within the external shock model. However the spectrum obtained is too high energy to be interpreted as synchrotron emission. Moreover it suggests the existence of another emission peak in the SED, in addition to the synchrotron peak around 10 keV, as it requires spectral hardening below 0.2 TeV despite the decreasing SED below 1 GeV. Determining the theoretical modeling parameters for the synchrotron peak, the other peak at higher energy can be explained naturally by the Synchrotron-Self Compton process (SSC). The temporal evolution of the spectrum also suggests the decrease of the peak energy with time, consistent with the deceleration nature of the external shock. In the following, I explain the discussion above in detail. The physical parameters mentioned here are explained in more detail in section G.2.

Connection to the afterglow indicated by the light curve

Figure 5.7 shows the light curves measured in various frequencies. In the afterglow epoch, the light curve of MAGIC (green data points) has the extended duration and the smooth, power-law temporal decay of the radiation. It suggests a direct connection between the TeV emission and the broadband afterglow emission. As discussed later, MAGIC energy range is closely related to X-ray (the energy range 1- 10 keV) in theoretical scenarios. The indices of the power-law decay of flux $F \propto t^{-\alpha}$ are $\alpha_{\rm X} \sim -1.36 \pm 0.02$ and $\alpha_{\rm TeV} \sim -1.51 \pm 0.04$, respectively.

Need of SSC interpretation

The SEDs of the radiation detected by MAGIC are shown in figure 5.8, where the whole duration of the emission detected by MAGIC is divided into five time intervals. For the first two time intervals, observations in the GeV and X-ray bands are also available. During the first time interval (68–110 s; blue data points and blue confidence regions), *Swift*-XRT, *Swift*-BAT and *Fermi*-GBM data show that the afterglow synchrotron component peaks in the X-ray band (around 10 keV). At higher energies, up to 1 GeV, the SED is a decreasing function of energy, as supported by the *Fermi*-LAT flux between 0.1 and 0.4 GeV. On the other hand, the MAGIC spectra above 0.2 TeV indicates the need for a spectral hardening between GeV energies and 0.2 TeV. This evidence is independent of the EBL model adopted to correct for the attenuation. This demonstrates that the newly discovered TeV radiation is not a simple extension of the known afterglow synchrotron emission, but a separate spectral component.

The most natural way to explain the TeV emission is through SSC in the external forward shock. The other possible mechanisms are proton synchrotron or the Inverse Compton in internal shock, but they are too inefficient to explain the data.

Reasonable values of the model parameters

For a reasonable theoretical modeling of the spectra, the physical parameters need be balanced among each other, especially for the bulk Lorentz factor Γ, the electron Lorentz factor γ, and the magnetic field B. From the peak energies of synchrotron $\nu_{\rm peak}^{Sync} \sim 10$keV and SSC $\nu_{\rm peak}^{SSC} \sim 100$GeV, eq.(G.2.21) tells that the SSC peak is

$$\begin{cases} \nu_{\rm peak}^{\rm SSC} \sim 2\gamma_e^2 \nu_{\rm peak}^{\rm Syn} & \text{Thomson regime} \\ \nu_{\rm peak}^{\rm SSC} \sim 2\gamma_e \Gamma m_e c^2 & \text{Klein-Nishina regime} \end{cases}$$

Therefore the electron Lorentz factor needs to be $\gamma_e \lessapprox 10^3$ in any case, which is relatively small in GRBs. This value is difficult to reconcile with the requirement from the synchrotron peak energy. From eq.(G.2.17), the peak of the observed synchrotron emission is

$$\begin{aligned} \varepsilon_{\rm peak}^{\rm Syn} &= \frac{1}{1+z}\frac{3h}{4\pi}\Gamma\gamma_{e,\rm peak}^2\frac{eB'}{m_e c} \\ &= \frac{1}{1.42}35\left(\frac{\Gamma}{20}\right)\left(\frac{B'}{10\rm G}\right)\left(\frac{\gamma_{e,\rm peak}}{10^5}\right)^2 \rm keV \end{aligned}$$

This relation yields $\Gamma B' > 10^5$ G, which requires both Γ and B' to be large. However if the bulk Lorentz factor Γ is larger than 150 (which is a necessary condition to avoid strong γ-γ opacity) it leads to another problem with the electron acceleration efficiency ϵ_e. It connects the bulk Lorentz factor in the shock Γ and the minimum Lorentz factor of the accelerated electron γ_m, via eq.(G.2.5),

$$\gamma_m = \epsilon_e \frac{p-2}{p-1}\frac{m_p}{m_e}\Gamma$$

Thus the large Γ and small γ_m require a small efficiency of the electron acceleration with $\epsilon_e < 0.05$. If B' is large and ϵ_e is small, the emission will be dominated by synchrotron and the SSC emission becomes inefficient. Therefore the low $\nu_{\rm peak}^{SSC}$ challenges the theoretical interpretation of the data.

The reasonable way to solve this problem is to consider the opacity by γ-γ absorption. As it happens in the EBL absorption (section G.3), two photons collide and produce e^+e^- pair. The target photons for γ-rays of energy $E_\gamma \sim 0.2$–1 TeV and for $\Gamma \sim 120$-150 have energies in the range 4-30 keV. This absorption plays a role in sharpening and softening the observed SSC spectra, and gives a non-negligible contribution to the radiative output.

Plausible result

Acceptable models of the broadband SED are obtained if the conditions at the source are the following. The initial kinetic energy of the blast wave is $E_k \simeq 3 \times 10^{53}$ erg (isotropic-equivalent). The electrons swept up from the external medium are efficiently injected into the acceleration process and carry a fraction $\epsilon_e \simeq 0.05$ - 0.15 of the energy dissipated at the shock. The acceleration mechanism produces an electron population characterised by a non-thermal

energy distribution, described by a power law with index $p \simeq 2.4$ - 2.6, an injection Lorentz factor of $\gamma_m \simeq (0.8 \text{ - } 2) \times 10^4$ and a maximum Lorentz factor of $\gamma_M \simeq 10^8$ (at about 100 s). The magnetic field behind the shock conveys a fraction $\epsilon_B \simeq (0.05 \text{ - } 1) \times 10^{-3}$ of the dissipated energy. At $t \sim 100$ s, corresponding to a distance from the central engine of $R \simeq (8 \text{ - } 20) \times 10^{16}$ cm, the density of the external medium is $n \sim 0.5$–5 cm^{-3} and the magnetic field strength is $B \simeq 0.5$ - 5 G.

The values of the parameters above fall within the range typically inferred from broadband (radio to GeV) studies of GRB afterglow emission. The magnetic field was efficiently amplified from values of a few μG, which are typical of the unshocked ambient medium, owing to plasma instabilities or other mechanisms. The relation of the efficiency $\epsilon_e \gg \epsilon_B$ is a necessary condition for the efficient production of SSC radiation.

This theoretical scenario does not require special conditions to be at work, which indicates that the SSC emission can occur in the usual framework of GRB afterglow. In other words, the SSC emission may be common among GRBs, despite the difficulty in detecting it with the current instrumentation.

TeV component consistent with the deceleration of the jet

Spectral softening is marginally observed throughout the duration of the GRB emission. This is difficult to attribute to γ-ray absorption. The absorption due to opacity should decrease with time because the radiation field becomes less dense due to deceleration and the lower density medium. Instead, it can be naturally explained by the change in emission due to well known deceleration nature of the jet. The deceleration of the jet causes the synchrotron peak shift towards the lower energies, and the SSC peak shifts correspondingly. Consequently, the spectral softening occurs as the relation between observed energy range and the SSC peak changes from lower energy side (harder) to higher energy side (softer).

Given that the emission occurs in the fast cooling regime, the peak shift should be rather fast[8], and it can be even the time scale of several tens of seconds. An example of the theoretical modelling in this scenario is shown in figure 5.9, in which first two time intervals of 68–110 s and 110–180 s, which are the brightest, are selected. From the modelling, the energy in the synchrotron and SSC component are estimated to be about 1.5×10^{52} erg and around 6.0×10^{51} erg respectively in the time interval 68–110 s, and about 1.3×10^{52} erg and around 5.4×10^{51} erg respectively in the time interval 110–180 s.

The time development of the TeV component is also consistent with the usual framework of GRB afterglow. This supports the insights that the SSC emission can occur in the usual framework of GRB afterglow.

The influence on our knowledge about GRBs

It has been more than 15 years for the MAGIC telescopes to finally detect the TeV emission from a GRB, although the observations have been performed on average about once per month. However, the TeV emission recently detected,

[8] In the slow-cooling regime, the SSC peak evolves to higher frequencies for a wind-like medium, and decreases very slowly ($\nu \propto t^{-1/4}$) for a constant-density medium (both in the Klein-Nishina and Thomson regimes). In the fast-cooling regime, the evolution is faster ($\nu \propto t^{-1/2} - t^{-9/4}$) depending on the medium and regime.

which can be explained as the SSC component, may be a process occurring in many GRBs, as the modeling does not require special conditions to account for the additional component. The detection had been possible this time because the distance to the GRB was relatively closer among GRBs, and the the observation conditions were relatively optimal.

The new SSC component accounts for substantial fraction of energy emitted in the GRB afterglow. The modeling tells that the total energies of emission are comparable between the synchrotron and SSC component. These results are anticipated to be a step towards a deeper understanding of the physics of GRBs and relativistic shock waves.

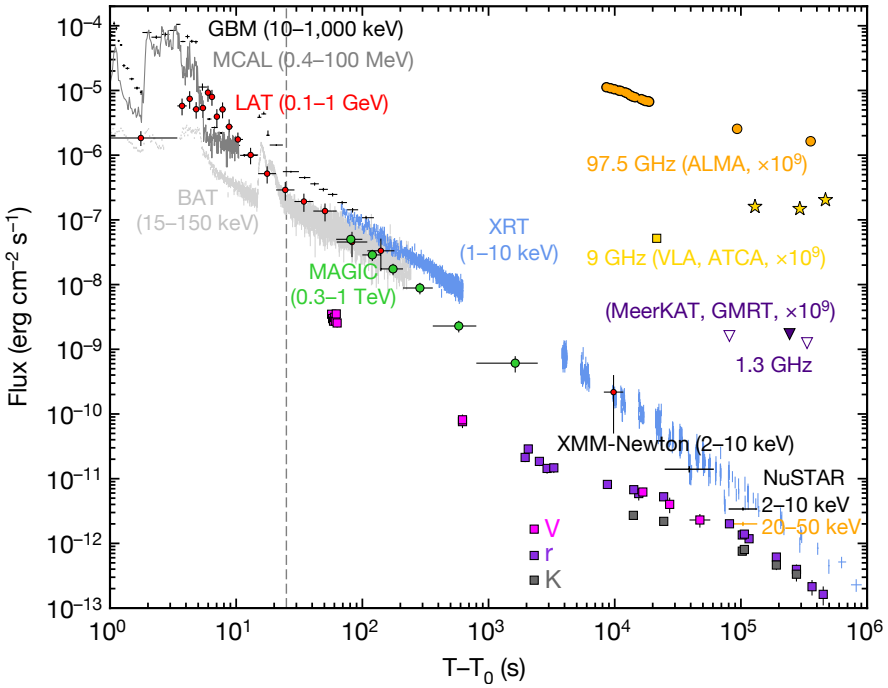


Figure 5.7: Multi-wavelength light curves of GRB 190114C

Energy flux at different wavelengths, from radio to γ-rays, versus time $t = T - T_0$. The light curve for the energy range 0.3–1 TeV (green circles) is compared with light curves at lower frequencies. Those for VLA (yellow square), ATCA (yellow stars), ALMA (orange circles), GMRT (purple filled triangle) and MeerKAT (purple open triangles) have been multiplied by 10^9 for clarity. The vertical dashed line marks approximately the end of the prompt-emission phase. For the data points, vertical bars show the 1σ errors on the flux, and horizontal bars represent the duration of the observation. The fluxes in the V, r and K filters (pink, purple and grey filled squares, respectively) have been corrected for extinction in the host and in our Galaxy; the contribution from the host galaxy has been subtracted. Figure extracted from [7].

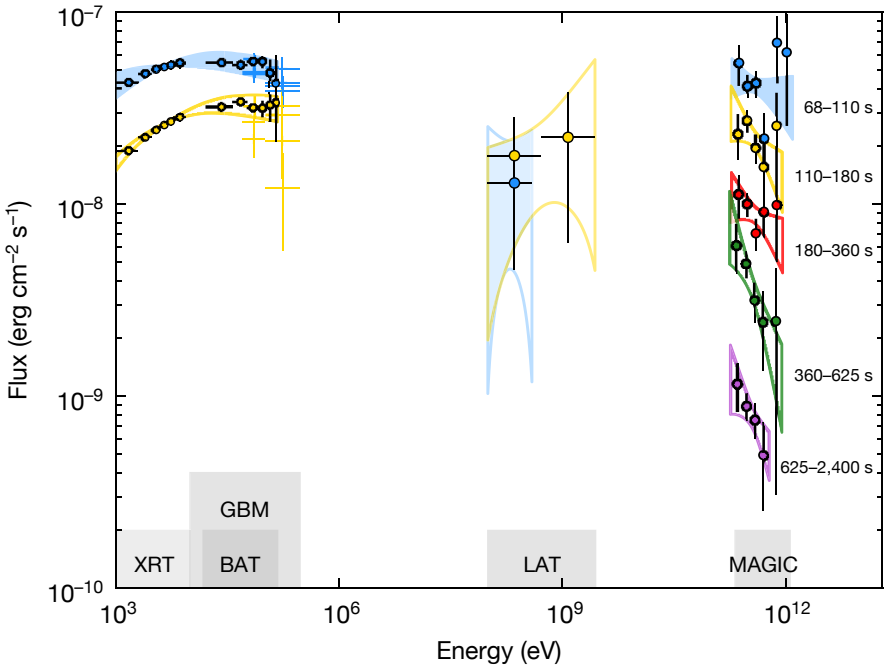


Figure 5.8: Multi-band spectra of GRB190114C in several time intervals.

Five time intervals are considered: 68–110 s (blue), 110–180 s (yellow), 180–360 s (red), 360–625 s (green) and 625-2,400 s (purple). MAGIC data points have been corrected for attenuation caused by the EBL. Data from other instruments (*Swift*-XRT, *Swift*- BAT, *Fermi*-GBM and *Fermi*-LAT) are shown for the first two time intervals. For each time interval, LAT contour regions are shown, limiting the energy to the range in which photons are detected. MAGIC and LAT contour regions are drawn from the 1σ error of their best-fit power-law functions. For *Swift* data, the regions show the 90% confidence contours for the joint fit for XRT and BAT, obtained by fitting a smoothly broken power law to the data. Filled regions are used for the first time interval (68–110 s). Figure extracted from [7].

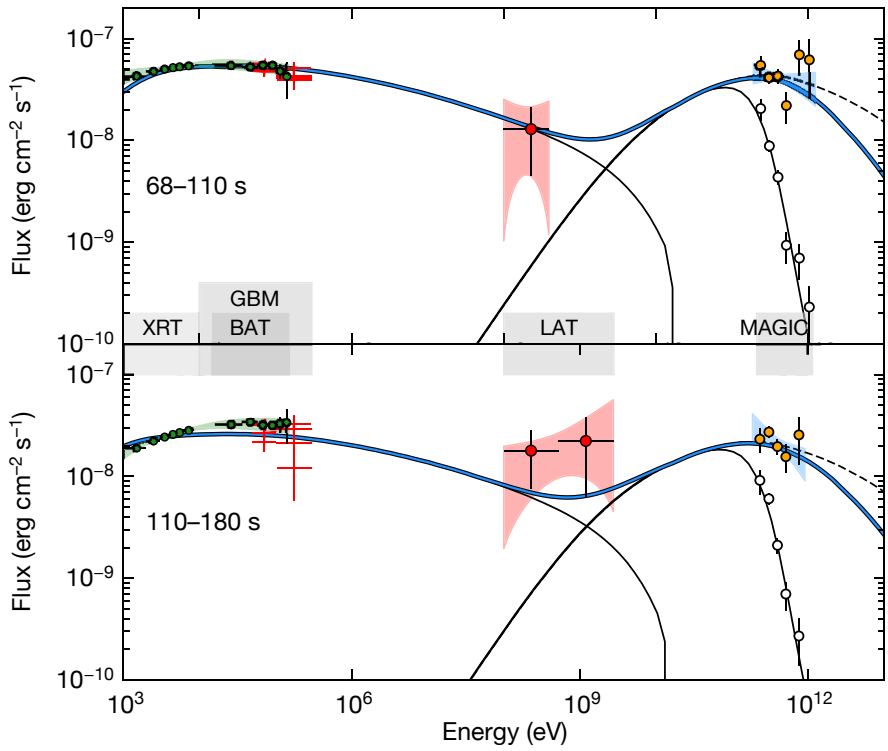


Figure 5.9: Theoretical modelling of the broadband spectra of GRB 190114C in the time intervals 68–110 s and 110–180 s.

The thick blue curve shows the theoretical model of the multi-band data in the synchrotron and SSC afterglow scenario. The thin solid lines show the synchrotron and SSC (observed spectrum) components of the model spectrum, including EBL attenuation. The dashed curve shows the SSC when internal $\gamma - \gamma$ opacity is neglected. The adopted parameters are: $s = 0$,$\epsilon_e = 0.07$, $\epsilon_B = 8 \times 10^{-5}$, $p = 2.6$, $n_0 = 0.5$ and $E_k = 8 \times 10^{53}$ erg.

The circles show the observed spectrum, and the instruments used are indicated by the colors of the circles and the energy ranges; green are *Swift*-XRT and *Swift*-BAT, red are *Fermi*-GBM and *Fermi*-LAT, and white and orange are by MAGIC. White circles show the observed MAGIC spectrum, that is, uncorrected for attenuation caused by the EBL. Orange circles show the spectrum corrected for attenuation. Contour regions and data points are as in figure 5.8. Figure extracted from [7].

Chapter 6

Summary

In this thesis, I report the development of a novel methodology of energy reconstruction for the very high energy (VHE, energy $\gtrsim$ 50 GeV) γ-ray events detected with the MAGIC telescopes. This new method substantially improves the energy reconstruction performance and extends the capabilities of the MAGIC telescopes to study γ-ray sources. This novel method for energy reconstruction has been validated and used for the first time on the observation data of the gamma-ray burst (GRB) GRB 190114C, the first GRB detected significantly at VHE. The data led to two publications in Nature reporting the spectral analysis results (Nature, vol.575, p455-458 and p459-463), and one in PRL reporting the test of Lorentz invariance violation (LIV) via event-wise analysis of the energies of γ-rays (Phys. Rev. Lett. 125, 021301).

The γ-rays bring information from the non-thermal particles in the Universe. Some of these particles are accelerated to energies millions of times higher than those that can be produced in Earth-based laboratories. Analyzing the energy distribution of γ-rays and their temporal evolution or spatial distribution, we can study the non-thermal emission processes occurring in the extraordinary cosmic accelerators, which is the key to the parent particle population and its temporal evolution. Additionally, the γ-rays can also be used to perform a wide range of fundamental physics investigations, like the study of the diffuse extragalactic background light and the inter-galactic magnetic field, the test of LIV, and the search for new particles beyond the Standard Model, like the elusive Dark Matter and the Axion. The accurate determination of the energy of the incoming γ-rays plays a crucial role in all these studies.

The Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) stereo telescope system is a worldwide leading instrument for γ-ray astronomy in the energy range from about 25 GeV to beyond 100 TeV. It is a system of two IACTs with 17 m-diameter parabolic reflector, which detects γ-rays through the images of Extensive Air Showers (EAS), visible via Cherenkov radiation. Before my PhD thesis, the standard energy estimation of the γ-ray was based on the Look Up Tables method (LUTs-Erec), which maps the energy of the γ-ray and the values from the variables derived from the EAS images. This has been the energy reconstruction method used by the MAGIC collaboration over the last decade.

During the last 3 years, I have developed a novel methodology of energy reconstruction for MAGIC observation data based on a machine learning algo-

rithm called Random Forest (RF-Erec). I carefully selected the input variables and defined the strategies to use them and to configure the RF, based on the physical insights on the EAS development and the telescope capability, to maximize performance as well as robustness and flexibility to diagnose possible systematic effects. The validation studies were done in various ways. I demonstrate that the RF-Erec works reliably on both the Monte-Carlo simulated γ-rays and the real γ-rays.

Compared to the LUTs-Erec, the RF-Erec improves the energy reconstruction performance in many aspects. The main merits are the following.

- **Improvement in energy resolution by a factor $\sim$2**

  The energy resolution is quantified by the width of a Gaussian fitted to the error distribution, which focuses on the majority of the distribution. For a wide range of the observable energies, the improvement of energy resolution reaches roughly factor $\gtrsim 2$, and the improvement is even larger for high Zenith distance (Zd) observations. The RF-Erec energy resolution is 20 % at 100 GeV, and 11 % above 1 TeV in low Zd, 20 % at 1 TeV and 13 % above 10 TeV in high Zd.

- **Reduction in the energy errors dispersion by a factor $\sim$3**

  The standard deviation of the error distribution takes into account the entire distribution including outliers, therefore its main difference from the energy resolution relates to the outliers of the distribution. I quantify this difference as tail spread, $(SD - \sigma)/\sigma$ where SD is the standard deviation and σ is the width of the fitted Gaussian. RF-Erec yields a narrower distribution with the better agreement with the fitted Gaussian shape, than LUTs-Erec, which shows long tails. Therefore the tail spread is reduced to less than half in low Zd, and even more in the other Zd ranges, with the maximum reduction in high Zd to around 10 % (i.e. $\sim$ 10 times better).

- **Decrease in the energy bias**

  The energy bias is quantified as the peak of a Gaussian fit to the error distribution. The bias from the RF-Erec is smaller for all energies and Zd ranges.

- **Reduction of the systematic uncertainties**

  I show that the usage of geometry-related parameters in the LUTs-Erec method is a dangerous source of systematic uncertainties. The effect becomes noticeable in the high Zd observation. To remove the systematic effect, I adopted a different strategy in the RF-Erec, and confirmed that it makes the systematic effect insignificant.

The better capabilities of RF-Erec compared to LUTs-Erec are visible in the migration matrix, implying that the spectral shape is less deformed by misreconstruction of energy in individual events, and that the correction of the deformation is easier. Therefore, RF-Erec gives better performance on spectral analysis. Moreover the energy estimation of each event becomes reliable, enabling a reliable spectral measurement even in situations with poor statistics, event-wise analysis like for Lorentz Invariance Violation (LIV) study, and

a search for anomalies in the spectral shape. The benefit is not only a better accuracy, but also a wider applicability, such as for observations at high Zenith distance, and morphological together with spectral studies.

In this thesis, I evaluate the actual effect of the performance improvement on spectrum reconstruction. One of the cases with the biggest improvement is on a high Zd observation of a source with very steep power-law γ-ray spectrum. Using MC γ-rays, I show that, in these situations, the RF-Erec provides a number of improvements, which I list below.

- **Reduction in the spillover range**

  Spillover is the energy mis-reconstruction, causing the events to be migrated to different estimated energies from the true energy. If the spillover is large, one may obtain γ-ray spectra with artificial spectral features that would lead to wrong scientific interpretations. These effects are particularly important at the high end of the γ-ray spectra, where a few extra high-energy photons could have dramatic consequences for studies related to the EBL attenuation, Lorentz invariance violation, or searches for Axion-like-particles. While spillover extends to at most factor of a few in RF-Erec, it extends to more than one order of magnitude in LUTs-Erec. In the high energy range, where there low statistics, the spillover to high energy can produce a substantial impact on spectral-analysis results.

- **Improvement in the bin purity**

  Bin purity is the fraction of γ-ray events in a given estimated energy bin, whose true energies are also within the same energy range. High bin purity is needed for accurate and reliable spectral determination. For a source with typical power-law index, RF-Erec improves the bin purity above $\sim$ 1 TeV by $\sim$50 % in low Zd observations, and by a factor $\sim$ 2 in high Zd observations. In the case of a γ-ray source with a steep power-law of index -5 that is observed at high Zd, the RF-Erec provides a bin purity that is more than one order of magnitude higher than LUTs-Erec for energies above 1 TeV, hence enabling energy measurements that would not have been possible otherwise.

- **Improvement in the determination of the spectral shape**

  The effect listed above results in a better determination of the spectral shape. RF-Erec has a better ability than LUTs-Erec in estimating the slope and amplitude, and yields more reliable and consistent results among different spectral analysis strategies.

I have implemented the novel methodology into the MAGIC Analysis and Reconstruction Software (MARS). It is now provided as part of the standard data analysis framework, and the whole MAGIC collaboration can benefit from this methodology. After validation by the MAGIC software board, this novel strategy was approved to be used in scientific papers, and, because of the superior performance, it is now becoming the new standard energy reconstruction method in the MAGIC collaboration.

The first scientific application of the novel energy estimation was on the data from the MAGIC observations of GRB190114C. It was the first GRB detected significantly at VHE γ-rays after more than 15 years of dedicated searches with

the MAGIC telescopes, and more than 40 years with many ground-based γ-ray instruments. GRBs are the most luminous explosions in the universe, happening at cosmological distance. Since its serendipitous discovery in 1967, the GRBs were gradually understood as a merger of binary compact objects or a death of massive star. The detection of GRBs at VHE γ-rays had been long awaited as the last missing waveband in the electromagnetic spectrum. However the spectral shape is very steep due to the absorption by the Extragalactic Background Light (EBL), that reduces the γ-ray flux by factors of several hundreds at TeV energies. Moreover the observation of this event was performed at high Zd. Under these conditions, the previous method for energy reconstruction, LUTs-Erec, would not have been able to provide a reliable characterization of the VHE γ-ray spectral shape from GRB190114C.

The rich photon statistics enabled the characterization of the VHE spectra on timescales as short as 1 minute. The spectral analysis revealed that the spectrum has the steepest shape over one decade in energy (from 0.2 TeV to 2 TeV) that has been ever measured with MAGIC, and with any VHE γ-ray instrument to date. The broadband energy spectrum (SED) indicates that the TeV emission belongs to another component other than the classical synchrotron component. The broadband SED shows, for the first time, a peak at $\sim$ 100 GeV, in addition to the well known synchrotron peak at $\sim$ 10 keV. The new component is interpreted as SSC emission from the external forward shock of the GRB outflow, emerging from the same population of relativistic electrons as those which radiated the synchrotron photons. This interpretation is also supported by the same power-law temporal decay trend of the light curves at keV and TeV energies.

The detection and characterization of the TeV emission is a big milestone for understanding GRBs because it accounts for about half of the total energy output of the GRB afterglow. Moreover, our results indicate that, despite the technical difficulties in detecting TeV γ-rays from GRBs, the SSC emission may be a common process among GRBs, which implies the need to substantially update our knowledge about these extreme phenomena.

Appendix A

Derivation of bias-variance decomposition

$$
\begin{aligned}
MSE &= \mathbb{E}\left[(\hat{f}-y)^2\right] \\
&= \mathbb{E}\left[(\hat{f}-f-\epsilon)^2\right] \\
&= \mathbb{E}\left[\left(\hat{f}-\mathbb{E}\left[\hat{f}\right]+\mathbb{E}\left[\hat{f}\right]-f-\epsilon\right)^2\right] \\
&= \underbrace{\mathbb{E}\left[\left(\hat{f}-\mathbb{E}\left[\hat{f}\right]\right)^2\right]}_{(1)}+\underbrace{\mathbb{E}\left[\left(\mathbb{E}\left[\hat{f}\right]-f\right)^2\right]}_{(2)}+\underbrace{\mathbb{E}\left[\epsilon^2\right]}_{(3)} \\
&\quad +2\underbrace{\mathbb{E}\left[\left(\hat{f}-\mathbb{E}\left[\hat{f}\right]\right)\left(\mathbb{E}\left[\hat{f}\right]-f\right)\right]}_{(4)}+2\underbrace{\mathbb{E}\left[\epsilon\left(\hat{f}-\mathbb{E}\left[\hat{f}\right]\right)\right]}_{(5)}+2\underbrace{\mathbb{E}\left[\epsilon\left(\mathbb{E}\left[\hat{f}\right]-f\right)\right]}_{(6)}
\end{aligned}
$$

By the definition of Variance in eq.(4.1.9), the term (1) becomes $\mathbb{V}ar\left[\hat{f}\right]$. And by the definition of ϵ, the term $(3) = \mathbb{E}\left[\epsilon^2\right] = \sigma^2$.

We modify (2), (4) and (6) $\mathbb{E}\left[\hat{f}\right]$ and f are not random. And using the definition of the bias in eq.(4.1.8),

$$
\begin{aligned}
(2) &= \left(\mathbb{E}\left[\hat{f}\right]-f\right)^2 = \left(\mathbb{B}ias\left[\hat{f}\right]\right)^2 \\
(4) &= \left(\mathbb{E}\left[\hat{f}\right]-f\right)\mathbb{E}\left[\left(\hat{f}-\mathbb{E}\left[\hat{f}\right]\right)\right] = \left(\mathbb{E}\left[\hat{f}\right]-f\right)\underbrace{\left(\mathbb{E}\left[\hat{f}\right]-\mathbb{E}\left[\hat{f}\right]\right)}_{0} = 0 \\
(6) &= \left(\mathbb{E}\left[\hat{f}\right]-f\right)\underbrace{\mathbb{E}\left[\epsilon\right]}_{0} = 0
\end{aligned}
$$

For (5), taking into account that ϵ is independent from $\hat{f}$,

$$
(5) = \underbrace{\mathbb{E}\left[\epsilon\right]}_{0}\mathbb{E}\left[\left(\hat{f}-\mathbb{E}\left[\hat{f}\right]\right)\right] = 0
$$

Therefore we obtain

$$MSE = \left(\mathbb{B}ias\left[\hat{f}\right]\right)^2 + \mathbb{V}ar\left[\hat{f}\right] + \sigma^2$$

Appendix B

Implementation of RF-Erec in the MARS software package

Here I describe the implementation of the Random Forest energy reconstruction in the MAGIC Analysis and Reconstruction Software (MARS).

Adding time gradient correction functionality

As mentioned in section 4.4.6, the correction of TimeGradient helps the better Disp estimation. To implement the TimeGradient correction, I defined two kinds of classes, `MTmpPar` and `MCalcTmp`, and integrated them into the MARS software package. In the MARS framework, they work for each event via `MEvtLoop` class, which governs a loop process over all the events in given input files.

`MTmpPar` is a class to store the corrected TimeGradient as one of the parameters for an event, therefore it inherits from the basic class `MParContainer`. `MCalcTmp` is a class to calculate corrected TimeGradient and store it in the `MTmpPar`. To perform the calculation in a loop over the events, it inherits from the basic class `MTask`.

When objects of `MTmpPar` and `MCalcTmp` are registered into an object of `MEvtLoop`, the corrected TimeGradient for every event in given input files is calculated and stored. The latter tasks registered in the loop can refer to it.

Modification of RF classes

To realize the functionality of energy estimation, I have modified an existing RF class. The classes which can work as interface to the other classes are the following ones:

- `MRanForest` is the class to describe RF.
- `MRanForestGrow` is the interface for training the RF.
- `MRFGHSeparation` is the interface for applying RF as a classification, especially the *Hadronness* estimation.

- `MRFEnergyEst` is for regression value estimation and had been used only for Disp estimation for stereo observation.

The switching of the target value is added to the `MRFEnergyEst` class. When the target value is in logarithmic form, the estimated value will be computed back to the value as power of 10. The RMS of the estimation is also calculated correspondingly.

Adding training functionality to *Coach*

The functionality of training the RF is integrated into *Coach*, which is dedicated to the generation of all the estimators of event property.

All the functionality of *Coach* is written in `coach.cc` and the training of the energy RF is added there too. The program now takes a set of new additional arguments to switch on the RF energy estimation. When *Coach* is kicked together with the arguments, the training of RF for energy estimation is processed. The program loads the configuration file (called rc file), which is specified also as an argument when the program is kicked, and reads the configuration for the training listed as below.

- Input file path
- List of variables
- Condition of data cleaning (skip condition)
- Forest configuration parameters
- Output file name

The actual training process is done by creating instances of the `MRanForest` class and the `MRanForestGrow` class interacting with them. It is given to the loop process over the training events, loaded from input file. In the loop process, instances of the temporal variable classes, `MTmpPar` and `MCalcTmp`, are also given.

After the loop finishes, the RF file is ready and stored.

Adding reconstruction functionality to *Melibea*

All the functionality of giving the reconstructed event property is in *Melibea*, thus the energy reconstruction is also added there.

The main function of *Melibea* is written in `melibea.cc`, but it only configures arguments and kicks the actual tasks through an instance of `MJMelibea`. Thus `melibea.cc` was edited for configuration of the energy reconstruction, which reads switching the RF energy reconstruction (RF-Erec) and the RF file name via arguments. The actual energy reconstruction part was added into `MJMelibea`. Objects of `MEnergyEst` class (the class to store reconstructed energy) and `MRFEnergyEst` class are generated and they are given to the loop process over the input file. Before the loop process, `MRFEnergyEst` is configured in the way below

- RF file path

- Swap of variables (Not used for RF-Erec ver.5)
- Switching on the computations for logarithmic target value.

In the loop process, instances of the temporal variable classes, `MTmpPar` and `MCalcTmp`, are also given.

Appendix C

Contributions of variables to the energy reconstruction

C.1 The effect of Disp and DispImpact

As discussed in section 4.4.2, Disp and DispImpact are the important variables for the performance of energy reconstruction. In figure C.1 I show the contribution of Disp and DispImpact to accuracy of energy estimation by a control experiment. I only show low Zd observation for simplicity. To see the effect of them, energy was estimated to the same test data set by two different RF energy estimators; one includes Disp and DispImpact in the input variables, but the other omitted Disp and replaced DispImpact by ClassicalImpact. The result shows that the performance drops if the alternative parameters are used. The difference of the performance shows the contribution from these variables. In low Zd observation, the largest contribution is lowest energy, where bias is suppressed by several tens of percent. The smaller bias can be seen also in highest energy range. Resolution is also improved slightly from low energy to middle energy.

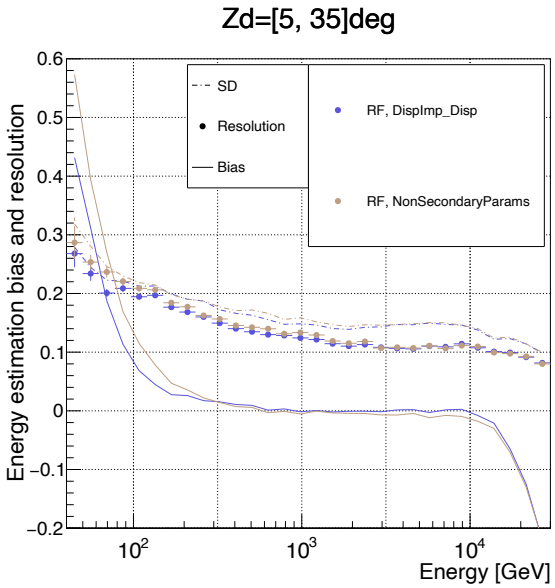


Figure C.1: The performance of the RFs with different treatments of Disp and Impact

The performances of the energy reconstruction by the two RFs are plotted. The bias (solid line), resolution (the filled circles), and the standard deviation (SD, dashed line) as a function of energy are drawn all together. The definition is explained in section 4.1.3. "DispImp_Disp" (blue) is with the inputs listed as v4 in table 4.2, but without Width and Length. The other, "NonSecondaryParams", is with the same inputs as "DispImp_Disp", except for Disp and Impact. Disp is omitted and Impact is ClassicalImpact, instead of DispImpact. The detail of the evaluation is explained in section 4.1.3.

Appendix D

Validation and performance evaluation of RF-Erec

D.1 Forest growing parameters

I explained in section 4.2 that RF does not require any elaborated tweak to optimize the performance, and in section 4.4.5 that it is also the case to the RF for the energy reconstruction. Thus it is unnecessary to care about the forest growing parameters in general for better performance, however they might be important when a user need to deal with the limitation on the computational resource. If the tree size is smaller or the node size is larger, the RF file size will be lighter. Therefore this section is intended also to show valid range of these forest growing parameters for lighter RF file size. To this end, the performance of RF-Erec is evaluated changing the number of trees and the node size.

Figure D.1 shows the result. The number of trees can be reduced to 50 without any significant degradation of the performance, and the minimum node size is still fine with 13.

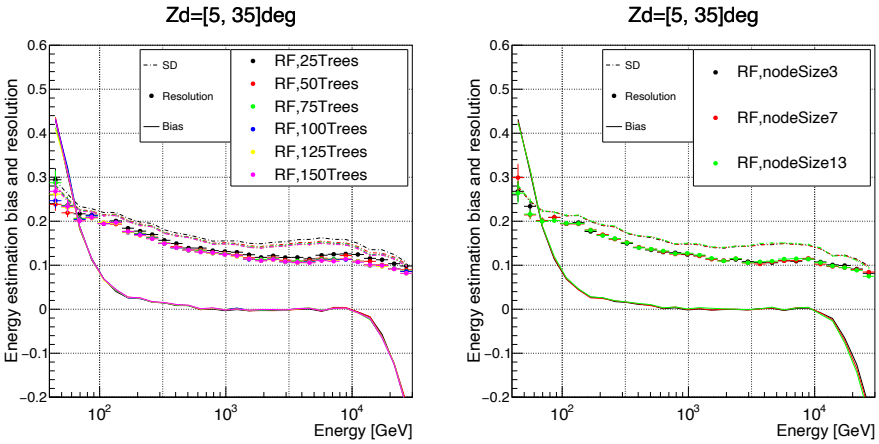


Figure D.1: The performances of different forest growing parameters

The performances of the energy reconstruction by the different settings of forest growing parameters. The left is for different number of trees, and the right is for different node size (the minimum number of events at the last node). The bias (solid line), resolution (the filled circles), and the standard deviation (SD, dashed line) as a function of energy are drawn all together. The definition is explained in section 4.1.3.

D.2 Applicability on morphological studies

The energy reconstruction works almost equivalently well, not only for the point source observation, but also for morphological study. In a morphological study, the analysis assumes that the source location in the field of view can be anywhere. Accordingly, the MC simulation of the γ-ray is performed as diffuse source throughout the field of view. The MAGIC collaboration provides the dedicated MC data sets of the γ-rays for morphological studies, as one of the standard analysis framework.

Here I show the basic performance of energy reconstruction for the morphological studies, using the diffuse source MC γ-ray data set. In the data set, the incoming directions are distributed uniformly within the circle around the camera center with radius of 2.5°, as shown in figure 3.4 and explained in figure 3.4.3. The date set is split into train and test samples, and the train set was used for generating RF and the test set was used for the evaluation.

Figure D.2 shows the performance properties on the diffuse source data, compared with the estimation on the standard ring wobble data. There is a discrepancy between them in bias in low energies, however this can be taken into account in the spectral unfolding. The energy resolution is very similar. The SD is also almost identical, except for the low energies. The slightly larger SD indicates that there are small fraction of outliers with large estimation error in the tail of the energy error distribution. However these outliers in the low energies would be negligible due to the uncertainty given by large amount of background events.

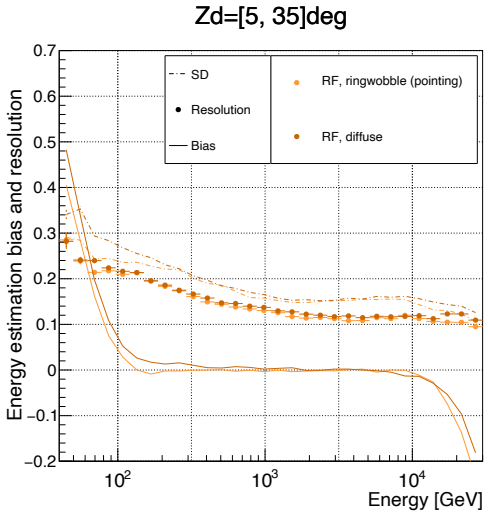


Figure D.2: Performance of the energy estimation on diffuse source

The performances of the energy reconstruction for the diffuse data and the ring wobble data (for pointing observation). Since the performance is evaluated for the morphological study, θ^2 cut is not applied. The bias (solid line), resolution (the filled circles), and the standard deviation (SD, dashed line) as a function of energy are drawn all together. The definition is explained in section 4.1.3.

D.3 RF-Erec-v3

Since the intermediate versions of RF-Erec were released, it may happen that a user needs to refer to the performance of a previous version. Here I describe the performance of RF-Erec-v3, because it was already used for some publications, like the two about GRB 190114C, that were published in Nature.

D.3.1 The basic properties of generalization error

Here I present the performance of RF-Erec-v3 (RFv3), comparing with the the latter versions (RFv4 and RFv5). The versions of RF-Erec are described in section 4.4.7. The most significant difference between RFv3 and the latter versions to cause the difference in the energy estimation performance is the treatment of Impact and Disp. In RFv3, the treatment of these variables are different in Zenith distance (Zd) ranges. In the low Zd range, Impact and Disp are used via swap strategy. In high Zd range, while Disp is used also via swap strategy, Impact is used by ClassicalImpact, to avoid the systematic effect reported in section E.3.

Figure D.3 is the basic performance, namely the bias, the resolution, and the standard deviation (SD) of estimation error, of RFv3, presented together with RFv4 and RFv5. The definitions of these properties of estimation performance are explained in section 4.1.3. They are shown in two different Zd ranges, which are Zd=[5, 35]deg (low Zd) and Zd=[55,65]deg (high Zd).

There is no significant difference between RFv4 and RFv5, while RFv3 shows some differences from the other versions. In low Zd range, RFv3 is better in

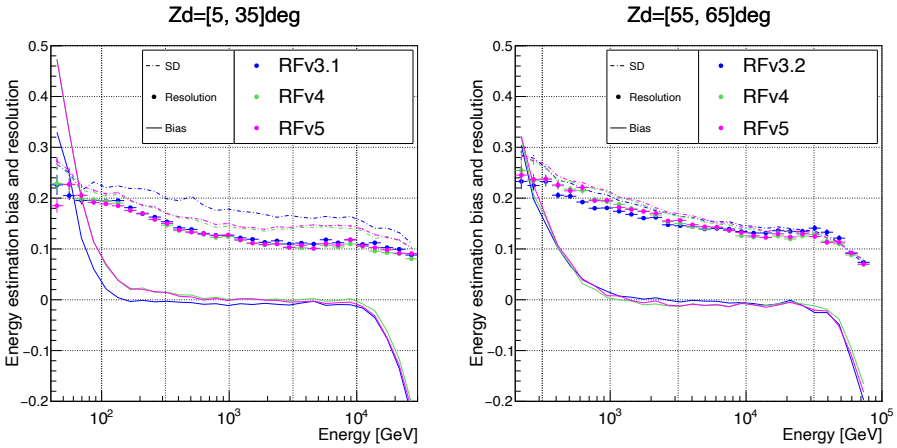


Figure D.3: The performance properties of different versions of RF-Erec

Performance of the different versions of RF-Erec energy estimators are shown, separately for different 2 ranges of Zenith distance (the ranges are shown on top of the plots). They are shown as a function of true energy (E_{true}). The solid lines depict the fitted bias, the filled circles show the fitted resolution, derived from the Gaussian fit, and the empty circles show the resolution as standard deviation of the distributions of the quantity $(E_{est} - E_{true})/E_{true}$, where E_{est} is the reconstructed energy. For more details of definition, see section 4.1.3.

bias below a few hundred GeV. However it is not as important as resolution or SD, because this can be taken into account in the spectral unfolding. RFv3 has worse SD in entire energy range, indicating wider spread of outliers. This is caused from the swapping strategy,in which the RF does not know the deviation in the estimated values from the true values. Therefore, the overall performance in low Zd is not as good as latter versions. As for high Zd, the performance look similar to each other.

D.3.2 Valid usage of RF-Erec-v3

It is also important to discuss the valid usage of RF-Erec-v3, because it contains the systematic effect which was the reason that it was overtaken by the latter versions. The validity of the performance of RF-Erec-v3 was investigated by drawing the spectrum of the Crab Nebula in the same way as section 4.5.4. In the investigation, there was no visible effect in the spectrum from low Zd observation. However, the spectrum from High Zd observation showed a visible feature, which was wiggling shape instead of smooth log-parabola curve. This anomaly is invisible as long as the Size cut and θ^2 efficiency cut remains strict[1]. The effect comes from the usage of Disp, and it is discussed more in detail in section E.3.2.

[1] For the data set I selected, the analysis with Size > 50 and θ^2 efficiency 85 showed the systematic effect, and it was invisible when Size > 80 and θ^2 efficiency 75

D.4 Energy threshold of Crab spectrum

I describe the details about the energy threshold in the Crab Nebula spectra for the validation of energy reconstruction, in figure 4.34. In the energy range close to the energy threshold for the detection of γ-rays, the collection area drops very quickly as the energy lowers and the number of events also drops despite the higher flux of γ-ray due to the power-law nature of the spectrum. Therefore the flux estimation suffers from heavy systematic effects in this range, and there needs to be an energy threshold for the spectrum.

To determine the energy threshold for energy spectra, I define the criteria in a histogram of excess events over estimated energy. The energy threshold is set to be the low edge of the lowest energy bin in the continuous bins with larger than 4 events and with the difference less than factor of 5 next to each other. The result is shown in figure D.4.

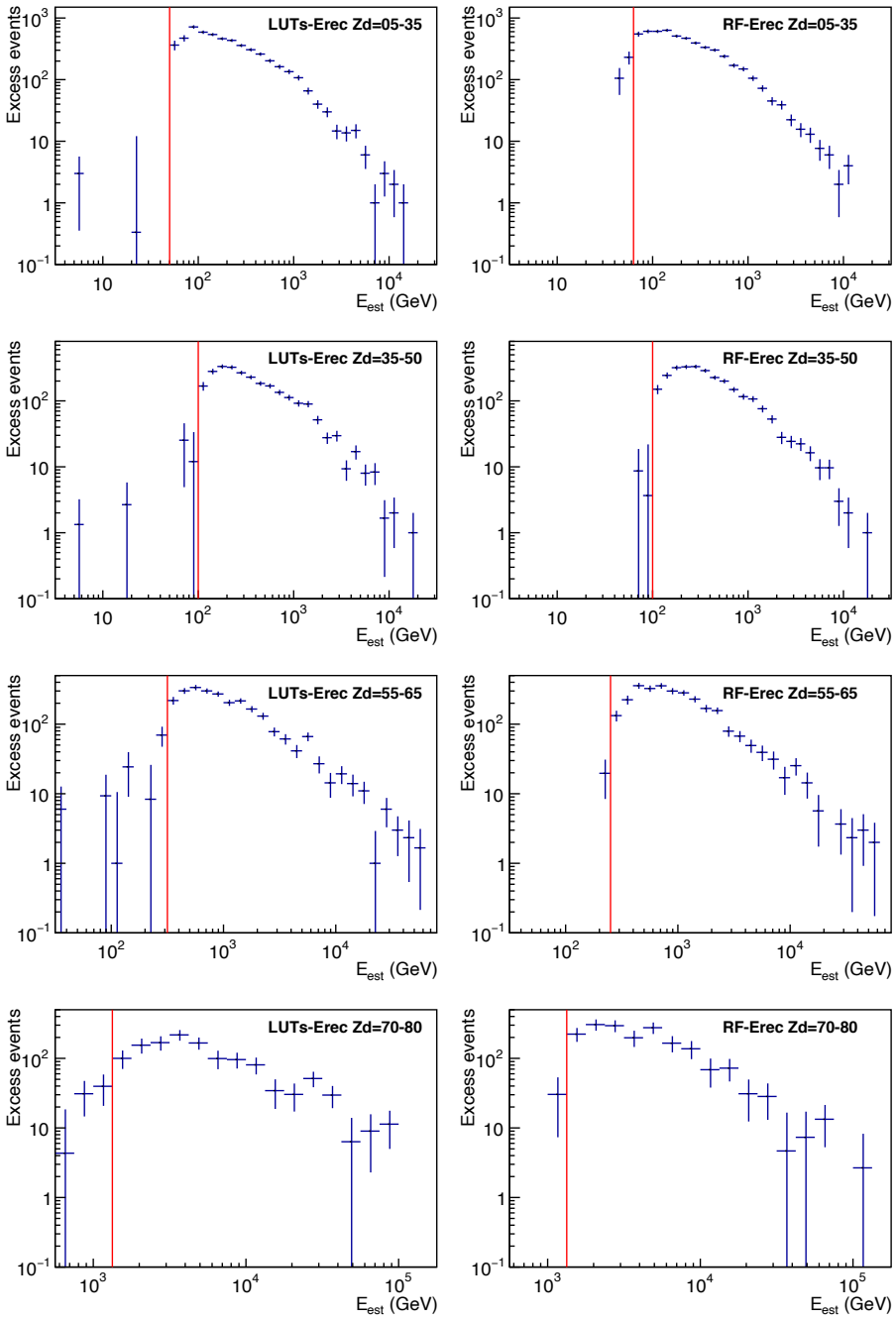


Figure D.4: The number of excess events in the analyses of Crab Nebula for validation

The histograms of the excess events are for the observation data of Crab Nebula taken in the Zenith distance (Zd) ranges indicated each panel. The strategies of energy reconstruction, RF-Erec or LUTs-Erec, are also indicated there. The vertical lines are the energy threshold determined by the criteria explained in the main text.

Appendix E

Investigations of the systematic uncertainties

E.1 RMS of prediction

In this section I discuss what the RMS of prediction indicates, and I conclude that prediction RMS is associated with uncertainty. I propose that an event with high prediction RMS can indicate shortage of statistics. I show a case study for applying this idea.

The definition of RMS of prediction is described in section 4.3.2. In the case of this study, the RMS of energy estimations from the individual trees is defined as EnergyRMS, described in section 4.4.4.

E.1.1 Understanding the prediction RMS

A RF is designed to give a prediction by combining different predictions by the trees it contains, and it is a part of designed feature that the RMS of the predictions by the trees has an amount larger than 0. The difference of the predictions by the trees comes from the randomization. Because of the randomization in selecting training data (bagging) and in selecting variable for split in construction of a tree, the iterative splitting of the nodes results in the populations at the last nodes, none of which are identical to any population at a last node in another tree, in ideal case (in which the decorrelation of trees works sufficiently by the randomization). The different populations lead to different predictive values. In the followings, I discuss more in detail about the meaning of RMS of prediction, focusing the role of randomization.

The prediction RMS and uncertainty

To work out the meaning conveyed by the prediction RMS, let us focus on the energy estimation by RF-Erec and define NERMS, which is the EnergyRMS normalised with the estimated energy, to treat the EnergyRMS equally over different energy scale.

$$\mathrm{NERMS} = \frac{\mathrm{EnergyRMS}}{E_{\mathrm{est}}} \tag{E.1.1}$$

Figure E.1 is the distribution of NERMS for the simulated γ-rays of different energies. In the two-dimensional histograms, the vertical axis is NERMS, and the horizontal axis is normalised energy error, $(E_{\mathrm{est}} - E_{\mathrm{true}})/E_{\mathrm{true}}$. The distribution tells that the NERMS is not correlated with the estimation error in general, moreover the distribution of NERMS peaks around 0.3 regardless of the estimation error.

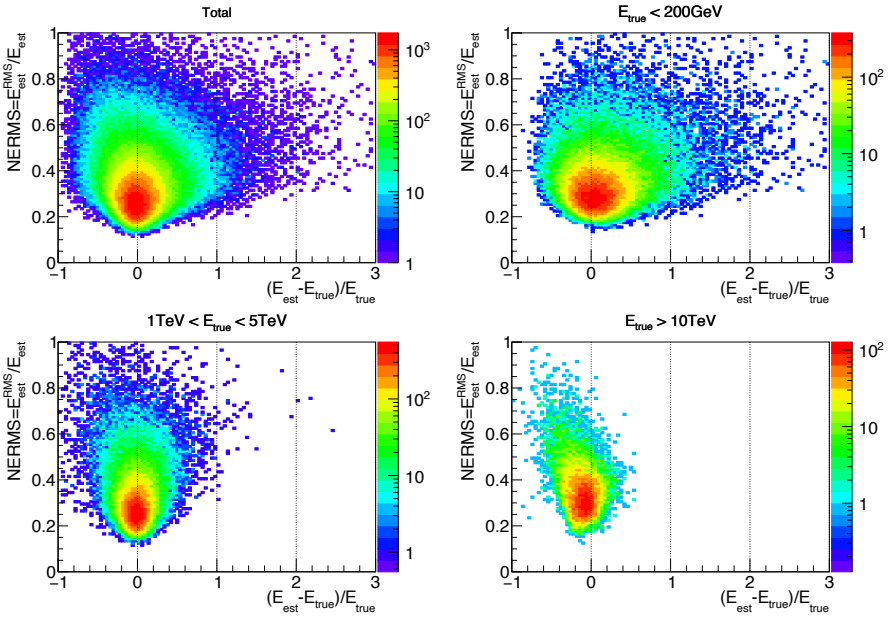


Figure E.1: NERMS vs normalised energy error

The relation between normalised EnergyRMS (NERMS) and the prediction error is presented for simulated γ-rays of different energies. The data is the test samples in high Zd range, dedicated to the GRB 190114C analysis (see section 5.3.1 for details). The vertical axis is NERMS, EnergyRMS normalised with estimated energy, and the horizontal axis is the estimation error normalised with the true energy.

However, NERMS is correlated with the width of the distribution of energy estimation error. Figure E.2 shows the standard deviation (SD) of normalised energy error as a function of NERMS, using the same data for figure E.1. The SD is positively correlated with NERMS almost in proportion especially in the range smaller than about 1, where there is enough statistics. It is reasonable to regard the SD of the normalised energy error to be uncertainty. Under this assumption, NERMS is correlated with uncertainty in almost proportional relation. Then the next question comes to if the NERMS is reasonable indication of uncertainty and how they are related. To understand this, let us begin with considering the purpose of the randomization, and continue with discussing separately on the randomization of samples and that of variables.

The purpose of randomization is to decorrelate the trees from each other so that the averaging of the predictions from the trees effectively mediates the bias in each tree. Ideally the trees can be decorrelated better if the subsamples for the training are different from each other, however it is unrealistic because simple

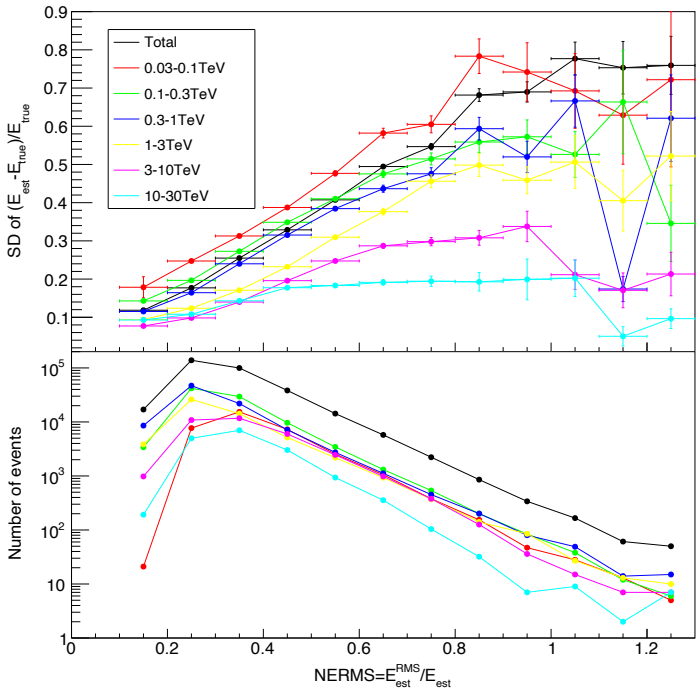


Figure E.2: Standard deviation of normalised energy error vs NERMS

The standard deviation (SD) of normalised energy error in the same simulated γ-rays, presented in figure E.1 is investigated. The different colors indicate the selected energy ranges.
Top panel: The standard deviation of normalised energy error is computed for the events in the ranges of NERMS.
Bottom panel: The number of events selected for the computation is shown.

separation of the training samples for individual trees significantly reduces the amount of subsamples. Nonetheless, to understand the ideal decorrelation, let us consider an extreme example, where approximately infinite amount of training data set is divided to give to training the individual trees, so that sufficient amount of subsamples can be given to each tree without duplication. A RF is a mapping between the input variables and the target value and the precision becomes high with rich statistics. However the mappings are different among the trees because the different sets of train samples are used. The difference must come from the fluctuation in input variables. Therefore, in this ideal case, the difference in the predictive values contains uncertainty originated from the range of variables indistinguishable from different energies, and the prediction RMS can be interpreted as uncertainty.

The randomization in RF works differently as an alternative way to the ideal decorrelation. By the bagging, namely the random subsample selection, a set of subsamples for a tree has a different population from that for the other trees. Although the populations of bagging subsamples are not exactly different because some events are shared, the bagging tries to decorrelate by an additional feature, which is to allow duplicated samples within the population, so that the

population itself is biased than the original population. This additional bias works for the tree to give a predictive value biased towards random direction. When the predictive values by all the trees in a RF are averaged, the additional biases from individual trees will also be averaged out.

A random variable selection in node split contributes to decorrelation too, by working as a random split, when none of the selected variables for a split is correlated with the target value. The best variable for a node split is selected from given number of trials, however, if the number of trials is sufficiently low compared to the number of weakly correlated variables, there are often the cases in which none of the important variables appears in the trials. In this case the split point of the samples over the variable is not related to the target value, and the split becomes random. The random split reduces the chance of the similar events to be clustered at the last node, which means adding an additional bias.

Above all, the randomizations plays a role to artificially add biases in random directions, which will be washed out in averaging the predictions from all the trees. This would be the reason to have NERMS larger than SD of normalised energy error, and it would be possible to introduce an uncertainty of energy estimation by using NERMS with the proportional constant. However, as shown in figure E.2, the relation changes with energy. Especially in the highest energy range, the predictive value is affected by the bias due to the truncation of energy range in training samples (see section E.2.4). Moreover, uncertainty does not mean actual energy error. As figure E.1 shows, there are outliers with high energy error even for the events with relatively low NERMS. Therefore, although there is a demand to remove high energy error events using NERMS, it is not reasonable to apply a cut on NERMS because large amount of events will be lost. Instead it is reasonable to take advantage of NERMS for pathological studies by investigating very high NERMS events, as the events with very high estimation error can be seen only in that range.

The prediction RMS and statistical shortage

If the the train data does not have enough statistics, the prediction RMS would become larger for the following reason. Training process of RF requires the train data to have enough statistics to dominate the last node by the events with the similar values of variables as well as similar values of target variable. The node size, namely the number of events at the last nodes, is as small as less than several down to three in general. If a location in the variable space lacks statistics, it will not be able to cover the corresponding last node, and the mixing with different population will happen. The mixing will cause estimation bias for a tree, as well as larger prediction RMS by the different prediction values tree by tree. The estimation bias will remain in RF, namely after the averaging over the trees, if the location in the parameters space with insufficient population is mixed mainly with that with abundant population.

In the next subsection I present a case study to take advantage of the relation between the prediction RMS, estimation bias and lack of statistics.

E.1.2 Case study in HZd data

I apply the insight on the prediction RMS to the investigation on the energy spectral analysis of GRB190114C with RF-Erec, presented in section 5.4. EnergyRMS is used in a pathological manner to investigate the reliability of every single event. In this study, the investigation focuses especially on the energy range around 1 TeV, where it becomes difficult to claim the significance of the spectral data point due to the poor statistics. In addition, it is also difficult to claim the highest energy detected. In this situation, the reliability of every single event becomes important.

Let us focus on the range of estimated energy from 900 GeV to 5 TeV. From the reconstructed spectrum of GRB190114C, the highest energy observable to MAGIC should exist in this range. The top panel in figure E.3 shows the distributions of normalised energy error of the simulated test data for different NERMS. The tail of the distribution in the large normalised energy error side is dominated by large NERMS events, presumably because of larger energy uncertainty for the larger NERMS events. However the fraction of events with large energy uncertainty is still small even if the events with large NERMS are extracted, and there needs to be another effort to identify if there is a systematic effect in the large normalised energy error.

Based on the insight on prediction RMS, it is possible to contain large NERMS events emerging from statistical shortage. And in this energy range, the one obvious tendency to make statistical shortage is the increase of Leakage events with energy, shown in figure E.4. The events with Leakage start to appear before 1 TeV, and the fraction continues to increase for higher energy, finally dominating more than half in 10 TeV. Since the non-zero Leakage events are very few around 1 TeV, they will let the last node to be contaminated with high energy events with similar Leakage. Consequently the constructed estimation value will be dragged to higher energy and the non-zero Leakage events around 1 TeV start to have systematic effect, which is positive energy error.

This diagnosis is confirmed by the lower panel of figure E.3, which is the distribution of normalised energy error in different NERMS, for the events without Leakage. Compared to the upper panel of the figure, non-zero Leakage events are removed, and the distribution is almost the same except for the high energy error side of the high NERMS distribution of NERMS > 0.45. And the distribution for NERMS > 0.45 does not have high energy error events any more than the distributions for lower NERMS, while the smaller NERMS ranges still have large energy error events.

The subspace of parameter space with statistical shortage would now be removed. In this way, NERMS would be of use indirectly for identifying the events with possibly higher energy error. And if there needs to be a rigorous test for the non-zero Leakage events, it would be reasonable to evaluate if the number of non-zero Leakage events in this energy range is consistent or not, by comparing the fraction of non-leakage events to all the excess events in this energy range in observation data, with that in MC data.

However, the ability to identify high energy error events by NERMS is limited. It is inevitable that there still remain events with high energy error with low NERMS. Figure E.5 shows the NERMS distributions for normal events and the events with high energy error, after non-zero Leakage events are removed. The distributions are not significantly different. Moreover, the NERMS of high

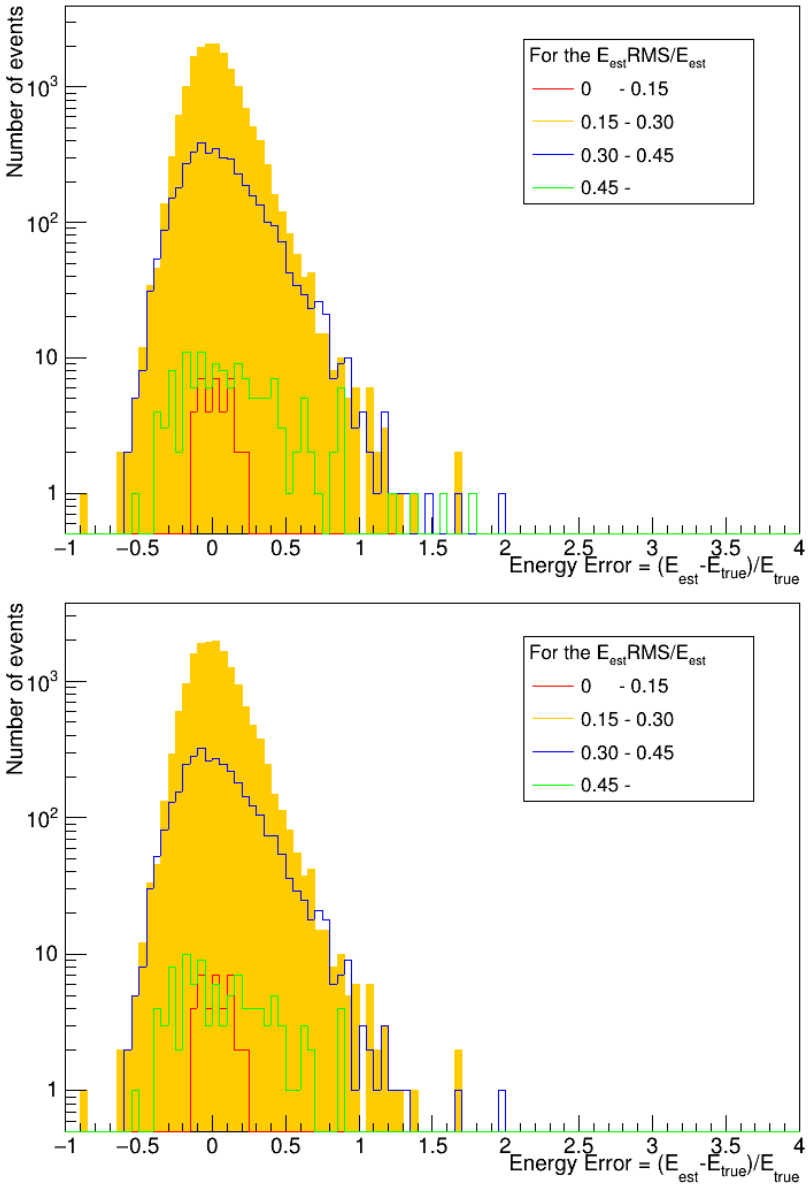


Figure E.3: The distribution of the normalised energy error for different NERMS

Each panel shows the energy error distribution for different NERMS ranges, shown in the legend. The data investigated is MC test data dedicated to the GRB 190114C analysis, which is in high Zd range with high NSB (see section 5.3.1 for details). The energy range is from 900 GeV to 5 TeV. The upper panel is for all the events, while the lower panel is for the events with no Leakage.

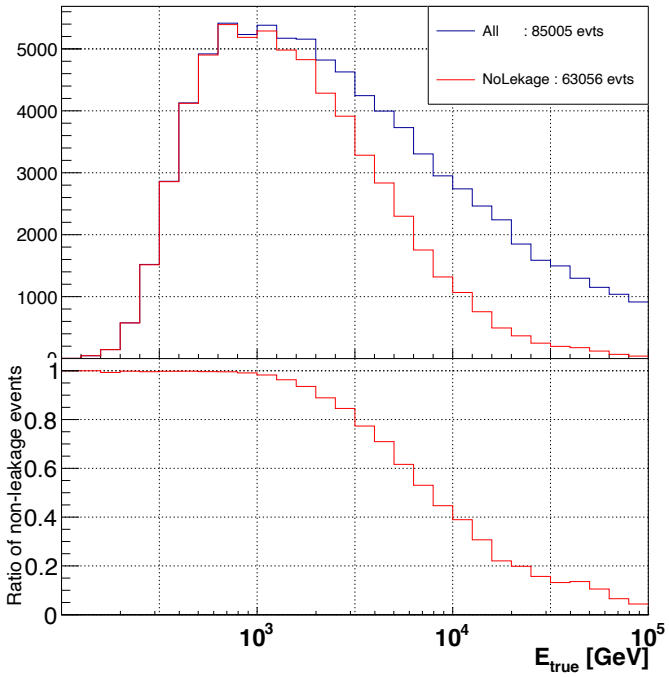


Figure E.4: The fraction of non-leakage events in the high Zenith distance MC data.

In this plot, the MC test data for figure E.3 is investigated. The histogram in the upper panel shows the number of events as a function of true energy, for all the events and for the events with no Leakage. The fraction of no Leakage events is shown in the lower panel.

energy error events is distributed only below 0.4. Therefore it is impossible to extract these high energy error events only by NERMS.

The reason why they have high energy error is revealed when the events are shown in the dimensions of Size and Impact. Since the energy is mostly determined from Size with the correction by Impact, an event with large energy error is likely to have wrong values of these variables. Especially Impact estimation can deviate from reality via the geometrical reconstruction of the event. To investigate such events, the distribution of events over Size and Impact is shown in figure E.6.

It clearly indicates that the events with high energy error are with higher error of Impact estimation. The top panels are the distributions in Size vs ImpactTrue (the true Impact measured in the simulation), in which the left panel is for M1 telescope and the right for M2. On the other hand, the lower panels are shown in Size vs ImpactEst (the estimated Impact via Disp method). The high energy error events are indicated as the scatter plots, on top of the two dimensional histograms of the normal events. Comparing the ImpactTrue and ImpactEst (the top panels and bottom panels), the estimated Impact of high energy error events is significantly higher than the true Impact. Moreover,

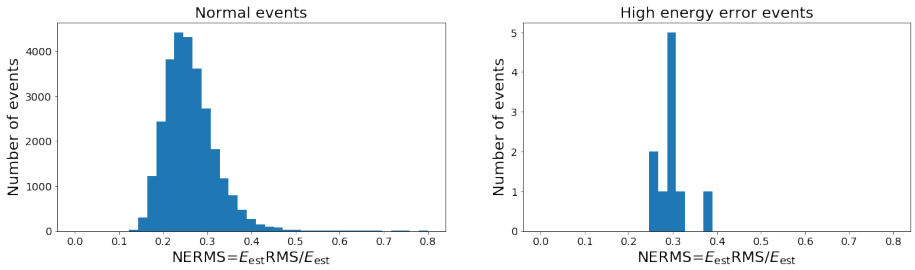


Figure E.5: Distributions of NERMS for normal events and high energy error events with no-Leakage

The distributions of NERMS for the MC test data for GRB 190114C, with estimated energy between 900 GeV and 5 TeV, Hadronness < 0.3 and $\theta^2 < 0.02$, after non-zero Leakage events are removed. The left panel is for all such events and the right panel is for high energy error. The additional cut applied is normalised energy error > 1.0.

most of them are deviated in the true Impact from the dense region of the distribution, while the estimated Impact is in the dense region. It is reasonable to regard them that they spilled over from low energy, due to the deviation of estimated Impact. The exceptional two events would be estimated to be even higher energy.

As the figures indicate, the events with high energy error can be with normal NERMS, if the energy estimation is presumably affected by high Impact error. The reason would be that they are just noise events in the variable space so that they do not form a group in the variable space. This example indicates that NERMS cannot be the general strategy to exclude the high energy error events.

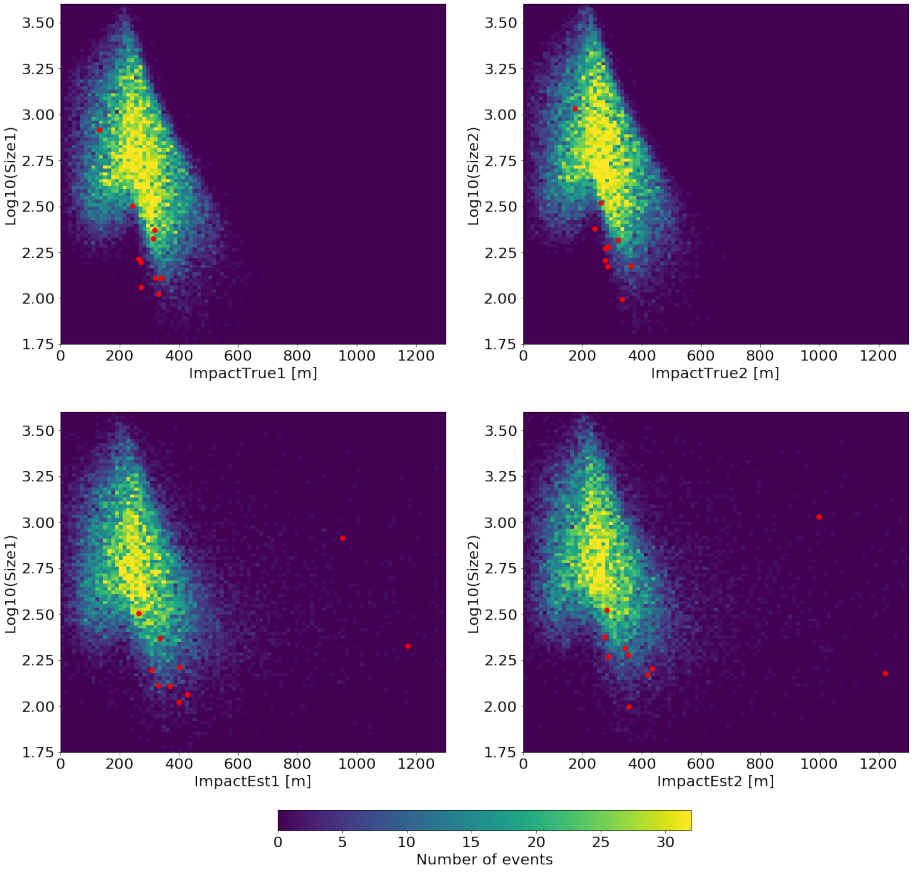


Figure E.6: Size vs. Impact, for normal events and high energy error events

The distribution of Size1 vs ImpactTrue1 (top left), Size2 vs ImpactTrue2 (top right), Size1 vs ImpactEst1 (bottom left), Size2 vs ImpactEst2 (bottom right) are shown. The events for the two-dimensional histograms are selected in the estimated energy between 900 GeV and 5 TeV, Hadronness < 0.3 and $\theta^2 < 0.02$. The scatter plots (red dots) overlaid indicate the events with normalised energy error larger than 1.0 but no Leakage.

E.2 The remaining (non-negligible) systematic uncertainties

Here I discuss the systematic uncertainties in energy reconstruction which are still difficult to remove. They arise in the specific regions of the parameter space. They are due to the suppression in statistics in the region, or the unbalanced confusion among the regions. In the former case, the difficulty is in accumulating statistics. In the latter case, it is in inevitable mixture in the parameter space. These effects can mostly be corrected by spectral unfolding process.

E.2.1 Sparse distribution of migration probability

Among the migration matrices presented in figure 4.35, the ones for Zd 55-65 deg have artifacts in the lowest-energy side of true energy; the estimated energy bins are not continuous for given true energy bin. Looking the sparse distribution in detail, the contents of the bins are too high to consider them as tail of the distribution. Moreover, the probability of migrating to the lowest estimated energy bin is not given by the lowest true energy bin.

These features are not reasonable to regard as having some physical origin, instead it would be due to very low number of events in low energy, as shown in figure E.7. In the lowest energy side the number of events is heavily suppressed by small collection area due to the energy threshold effect (drop of trigger efficiency and surviving rate after image cleaning). Consequently the number of events in the lowest energy is heavily suppressed. The problem is not only the number of events given for training, but also for the construction of the migration matrix. Therefore, one must be aware that the systematic effect can exist not only in the energy estimation, but also in the spectral unfolding.

The example of an artifact appearing in spectral unfolding is seen in section 4.6.4. In case of LUTs-Erec, the poor energy resolution results in the wide distribution of energy migration, and the effect is clearly visible (more details are described in section E.3.1). On the other hand, RF-Erec has good energy resolution and tight distribution of energy migration, thus it is less probable to create visible artifact in the analysis.

To prevent from the systematic effect as sparse distribution of the migration probabilities, there need to be larger statistics to construct smooth distribution of migration matrix. Since LUTs-Erec is heavily affected by this problem, it needs significantly more events than RF-Erec.

E.2.2 Bias in the lowest energy

As shown in figure 4.36, the bias of energy estimation becomes large in the lowest energy side. The reason is generally attributed to the threshold effect, in which the low energy events can be triggered and survive until the flux estimation, only for those which resemble a typical event with higher energy. However, there is another certain reason.

The distortion of energy distribution of train samples, coming from the drop of collection area in the low energy side, is also likely to be the source of the bias. In the training of RF, the estimated energy is determined as geometrical mean of events in the last node, because the MC simulation data for train and

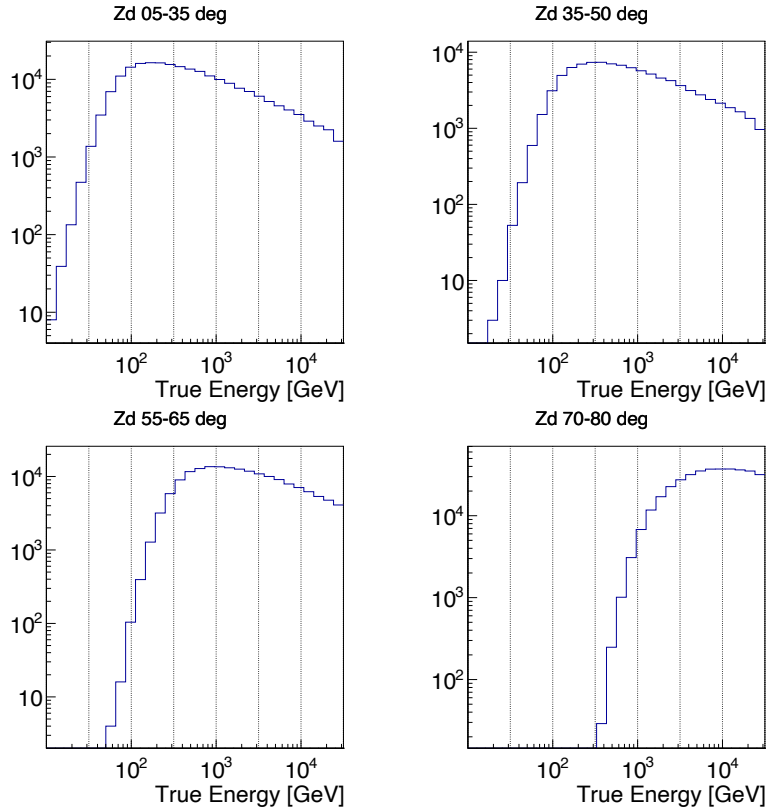


Figure E.7: Energy distributions for simulated γ-rays of different Zd ranges

These are the actual energy distribution for the samples given to the training of RF-Erec. The original energy distributions in the MC simulation of γ-rays is in power law, however it is heavily distorted in the lowest energy range due to the drop of the collection area. The distortion tendency coincides with the rise of energy bias at low energy side shown in figure 4.36.

test samples are generated in power law over energy. However, the distribution of γ-rays is exact power-law only at the stage when γ-rays are generated. Before the events are given to train an energy estimator, they must be processed via the simulation of the telescope triggering, the signal extraction and the image cleaning, down to the event parametrization. In low energy side, the probability of surviving these steps quickly drops towards energy threshold and the actual events available for the training is heavily distorted, as shown in figure E.7. The distortion affects in a way that the geometrical mean of the events at the last node tends to shift the estimated energy upwards, because the events to take average is always dominated by higher energy events.

This is consistent with the trends of the energy distribution and the bias. The bias, shown in figure 4.36, start to increase where the energy distribution departs from the power-law distribution.

E.2.3 Bias in the middle energy range and at high Zenith distances

The performance of energy estimation is not uniform with Zenith distance (Zd), and it becomes a source of systematic effect as bias. The bias appears in the middle energy range, where the bias keeps flat. While the bias in the middle energy range is almost zero for lower Zd side in the applicable range of RF, the bias becomes negative by a few to several % for higher Zd side.

The effect comes from the confusion of the events among different Zd. As explained in section 4.4.1, the energy estimation is mainly based on Size, and the other variables work as corrections, mainly to take into account the geometrical configuration of the shower with respect to the telescope. Therefore, it is likely to mix the events with similar Size but in different Zd.

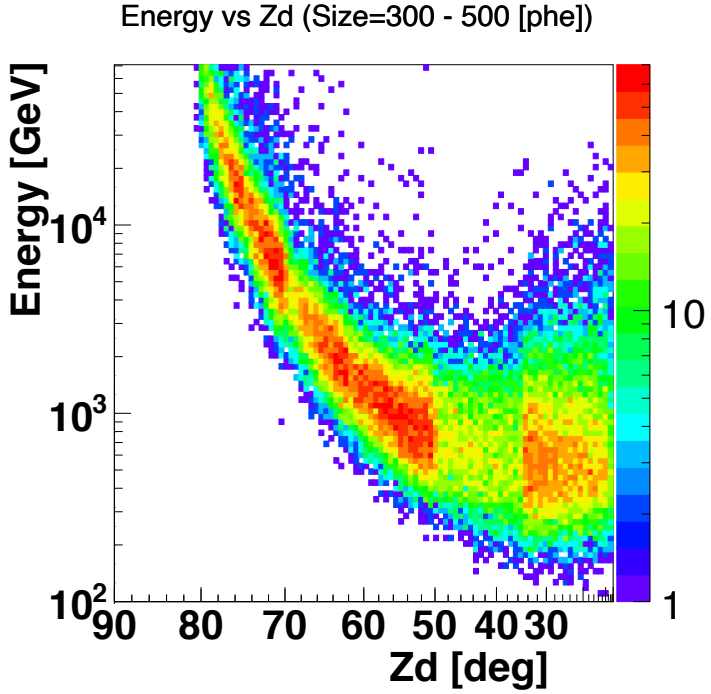


Figure E.8: Energy vs. Zd for the events with the same range of Size

The distribution is drawn from different mass productions of MC data for different Zd ranges. Thus the number of events generated is not uniform with Zd. However, the figure tends to show the typical energy for different Zd with similar Size.

Figure E.8 shows the event distribution of simulated events which have Size for M1 between 300 and 500 phe. For the events with the same Size, the energy is higher when the Zd is higher. The gradient of energy with Zd is steeper in higher Zd. Therefore, if a last node for typically lower Zd is contaminated with one with higher Zd, the estimated energy will be determined as higher value, on the other hand, the node for higher Zd can be contaminated with lower Zd events, lowering the estimated energy.

The contamination becomes likely if the parameter space is difficult for the energy estimation. As shown in figure 4.36, the higher the Zd, the worse the energy resolution. Thus the confusion is likely for higher Zd range and this would be the reason why the lower Zd range has less bias than higher Zd.

E.2.4 Bias in the highest energy range

As shown in figure 4.36, the bias of energy estimation becomes large in negative direction in the highest energy side. This originates from the truncation of the energy distribution of the MC simulation at highest energy.

When the training process of RF determines the estimated energy, the predictive values for a tree are calculated for the last nodes, from the average (geometrical mean) of the true energies for the events in the corresponding nodes. Therefore the estimated energy cannot be larger than the maximum energy of the training samples, and the estimated energy starts to be lowered when the distribution at the last node lacks the "proper" event at higher energy side; the information about the EAS is limited and the confusion among different energy cannot be removed, therefore the confusion in the parameter space must be reproduced in the training process.

The energy range in interest should be free from the bias, by extending the energy range of the simulation to enough high energy. The suitable maximum energy to simulate would be around double of the maximum energy in interest.

E.3 The systematic uncertainties from swap strategy

As introduced in section 4.4.2, the swap strategy in training an estimator is to use the true value available in the simulation instead of the one from the measured parameters of the shower event. However, this strategy is vulnerable to noise and can easily produce visible artifacts. The accuracy of estimation of these variables is limited, and systematic effects will appear when the estimated value fluctuates significantly more than the noise level of the true values in the train samples.

In this section, the detail of the effect is discussed presenting examples as well as explaining the probable mechanisms.

E.3.1 Swapping Impact

The swap strategy for Impact is used not only in RF-Erec-v2, the intermediate version of RF-Erec, but also in LUTs-Erec. Although the appearance of the systematic effect is different between them, they can be explained consistently that the estimators did not learn other than what was provided; they did not learn the noise.

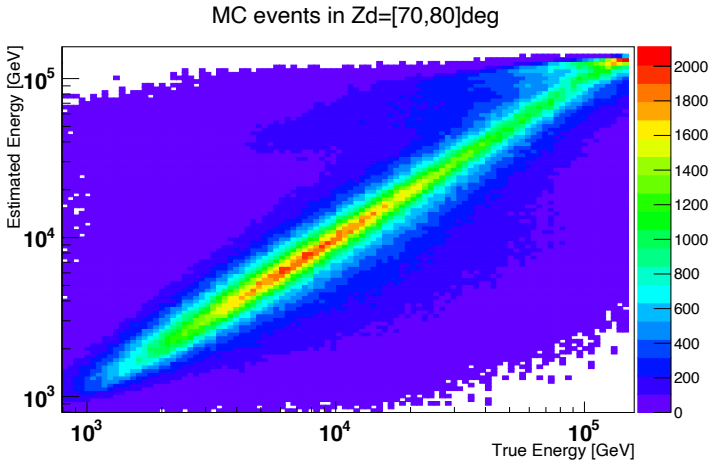
Noise of Impact

The error of Impact is interpreted as the error of Impact point. Since the Impact point is derived tracing the shower axis from incoming direction through the shower maximum to the ground, the error of incoming direction affects the direction of the vector towards the Impact point and results in the error of the Impact point. In case of low Zd, assuming the sufficient error of the direction reconstruction as $\Delta\theta \sim 0.2^\circ$ for the θ^2 cut and typical MaxHeight as ~ 10 km would make ~ 40 m of the error of Impact. The error dramatically increases as the Zd increases, because the shower develops at similar height and the distance to the shower from the telescopes becomes longer. In case $Zd = 80^\circ$, the distance to the shower would become possible to be even larger than 50 km. It would make ~ 200 m of the error of Impact.

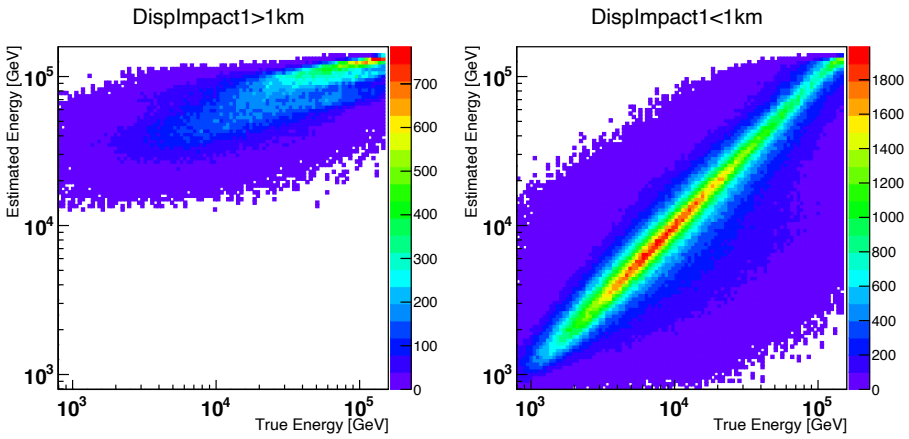
Swap Impact in RF-Erec-v2

In case of RF-Erec with swapping Impact strategy (RF-Erec-v2), the effect is clearly visible in the two dimensional distribution over estimated energy vs true energy, shown in the top panel of figure E.9. The distribution is not only around the diagonal line of $E_{\mathrm{est}} = E_{\mathrm{true}}$, but also around the line with constant E_{est}, close to the highest value of E_{true} and regardless of E_{true}. This effect can be explained by the noise of Impact in the following way.

An EAS at too large distance from the telescopes cannot be triggered, therefore there is a maximum Impact value set in a MC simulation, above which the trigger rate is negligible thus not simulated. In a high Zd range, the noise of Impact become too large that there become significant amount of event with deviation of Impact exceeding even far outside the range of the maximum Impact as shown in figure E.10.



(a) All the events



(b) DispImpact larger(left) and smaller(right) than 1km

Figure E.9: Systematic effects in RF-Erec-v2 from swapping Impact

The distribution of the MC simulation data with the Zenith distance in 70- 80 deg is drawn in estimated energy vs. true energy. Top panel is for all the events, bottom left is for the events with DispImpact $>$ 1 km, and bottom right is for the events with DispImpact $<$ 1 km. In the distribution for all the events, the branch can be seen in addition to the diagonal band, and the bottom panels show that it comes from large Impact.

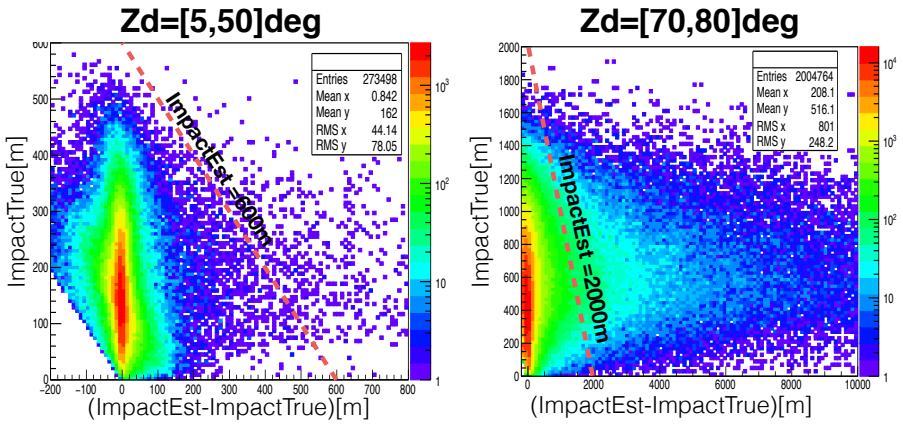


Figure E.10: True Impact vs. estimation error of Impact in two Zd ranges

The distributions of the MC γ-rays generated in two different Zd ranges. The horizontal axis is the estimation error of DispImpact. The error should be 0, no matter what the true Impact (the vertical axis) is. The distributions show accumulation around 0, however the distributions have tails and it is wider for higher Zd. The red dashed lines show the border of maximum Impact simulated, and the right side of the border indicates that the estimated Impact is larger than the maximum Impact.

Due to TrueImpact used for the training, Impact becomes strong clue to determine the energy. On the other hand, the energy is determined from estimated Impact, and it ranges to larger than the largest value of the TrueImpact. The events with the estimated Impact larger than the possible range of TrueImpact will be forced to refer to the relation between the maximum Impact and the energy. Since the maximum possible distance to be triggered enlarges as the energy becomes higher, the events with the largest TrueImpact exist only with the highest energies. This causes the saturated energy estimation, and creates the flat distribution of E_{est} around the maximum E_{true}. This view is confirmed by dividing the distribution by the estimated Impact, shown in the bottom panels of figure E.9. When the distribution is divided at 1 km of the estimated Impact, the two clusters are clearly separated.

Furthermore, the detailed shape, namely the branches in the saturated E_{est} distribution, also reflects the functionality of the RF. In the RF energy estimation, the estimated energy is determined as the average of the estimations by the trees. The trees are grown in different way, therefore Impact is not always the most influential variable, and the Influence of the estimated Impact varies event by event and tree by tree. Consequently the difference of the influences create the different numbers of trees to give the saturated energy changes, and the branches are created.

Swap Impact in LUTs-Erec

This effect is the same sort as described for RF-Erec in section E.2.1, however the effect described here for LUTs-Erec is more significant due to the swap

strategy for Impact.

In LUTs-Erec, the effect appears in two stages. The first stage, which is not yet a systematic effect, is the very wide distribution of the estimation error. As can be seen in figure 4.35, the tail of the estimation error distribution is significantly wider than RF-Erec, and it is even wider in higher Zenith distance (Zd) range. It comes mainly from the overlooked-noise of Impact; the Look-Up Tables (LUTs) of energy estimation are constructed based on the relation between the true energy and the true Impact, thus the error of the estimated Impact directly contributes to the error of estimated energy. The fluctuation of Impact is larger as the Zd is higher, thus the outlier of the energy estimation error is wider.

The second stage comes as the systematic effect in the migration matrix. The migration matrix requires sufficient statistics to reproduce the real probability distribution of migration. When the distribution of energy estimation error is wider, the range of possible estimated energy to a given true energy becomes also wider, thus the number of test samples, i.e. the number of samples to evaluate the energy migration, needs to be increased. However it is limited, and the migration matrix becomes different from the truth. As the actual effect, the probabilities distribute sparsely over energy with gaps, although each probability is high.

The visible effect can be seen clearly in the migration matrix for the Zd from 55 to 65 deg, shown in figure 4.35. The true energy around 100 GeV has the probability to migrate to around 5 TeV, as an isolated bin with high probability above a few tens of percent. As this example shows, the effect is more serious in the lowest energy, and the reason is related to Impact.

The low energy events cannot be triggered when the event is with large Impact, because the low energy events are too dim to be triggered at large distance. Therefore, if TrueImpact is used for training, the estimator learns the events with large Impact is related only to high energy. However, the estimated Impact, DispImpact can be largely mis-reconstructed in high Zenith distance, because the shower maximum becomes much distant from the telescopes and the impact point reconstruction will suffer from the large fluctuation. Thus the estimated energy becomes very high, especially for a low energy event with highly up-fluctuated DispImpact.

It is very unlikely to overcome the problem of sparse distribution in the lowest energy. Close to the energy threshold, the number of events to survive the trigger, image cleaning and analysis cut is very small, however the fluctuation of the energy estimation is much larger than low Zenith distance case. The wide distribution of impact parameters cannot be covered by very small amount of surviving events.

Above all, the LUTs-Erec cannot avoid the systematic effect in unfolding process, for a data with high Zd range observation. The detail of the visible effect on the spectral analysis is also discussed in section 4.6.

E.3.2 Swapping Disp

The swap strategy for Disp is used in RF-Erec-v3, the intermediate versions of RF-Erec. The systematic effect from this strategy is not as obvious as the other effects that I show. It is not visible in a migration matrix nor the basic performance. However it becomes significantly visible in a spectral analysis at

high Zenith distance (Zd) range. Firstly I show the control experiment drawing the SED of Crab Nebula by high Zd observation, to reveal the existence of the systematic effect and it is caused by the swapping Disp. Secondly I give the explanation on the mechanism of the effect.

Identifying the systematic effect by Disp from SED comparison

The systematic effect becomes noticeable when the spectrum for the same data is drawn by different energy estimators generated by different configurations of the input variables. Figure E.11 are the spectra of the same Crab Nebula observation data analysed in the same way, but by different energy reconstructions. The data was taken at the Zenith distance between 55 deg and 65 deg for 3 hours of total observation time. The energy reconstructions are based on the different energy estimators, which are all RF but different choices of input variables. They are based on RF-Erec-v2, and the input variables are changed, as the names indicate; "SwImp" indicates that SwapImpact is used, "ClImp" indicates that SwapImpact is replaced by ClassicalImpact, and "DispImp" indicates that SwapImpact is replaced by DispImpact (nested use of DispImpact). RF-Erec-v2 does not adopt any Disp, however, "Disp" indicates that Disp is added (nested use of Disp), and "SwDisp" indicates that SwapDisp is added. "L" indicates that Length is added.

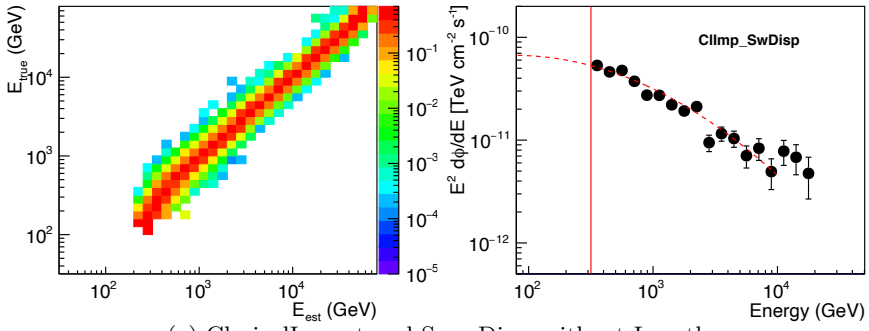
Comparing the 14 spectra, some show wiggled feature in the same way. They all adopt "SwDisp", on the other hand, those which do not adopt "SwDisp" do not show the feature. Therefore it is certainly better not to adopt SwapDisp strategy. It is also notable that the migration matrix is worse when input variables contain "SwImp", the SwapImpact. Therefore Swap strategy should be avoided.

Figure E.12 is the comparison of the basic performances of the 14 energy estimators. Among the estimators with safe performance, the "DispImp_Disp_L" shows the best performance. Therefore, the nested use of Impact and Disp is the best way as the input variables for the energy estimator. And Length contributes to the performance.

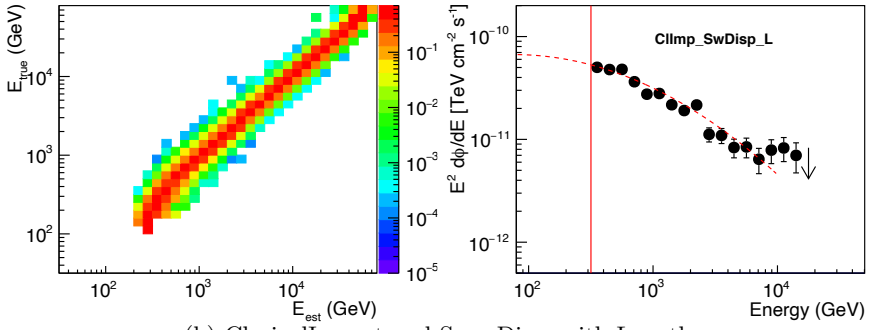
The mechanism of systematic effect from swapping Disp

As presented above, the systematic effect can be seen in the Crab Nebula spectrum drawn from high Zd observation data, and it comes certainly from the swap strategy for Disp. The mechanism of the systematic effect can be reasonably explained in the following way.

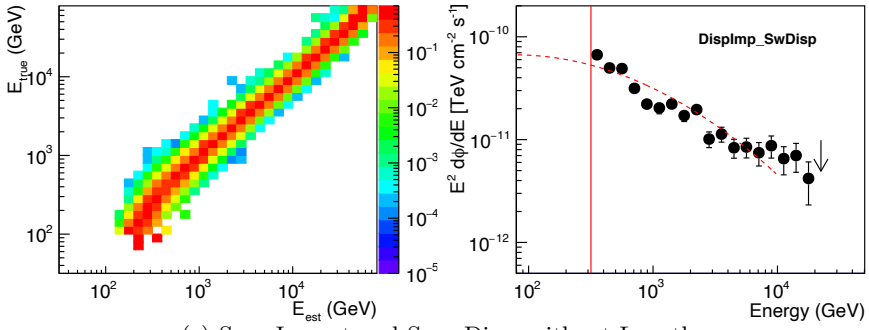
Since the systematic effect in the energy reconstruction appears in the use of SwapDisp, the source of the effect must be seen in Disp. To search for the relation to the energy, it is reasonable to decompose Disp into two geometrical factors, MaxHeight and Impact. As introduced in section 4.4.1, Disp is the angular distance between shower incoming direction and shower maximum direction, thus it is closely related to them. Especially, we can approximately attribute the systematic effect in Disp to the systematic effect in Impact because Impact is stronger information than MaxHeight. Therefore I compare the systematic difference between the TrueDisp and estimated Disp, in the relation with TrueImpact.



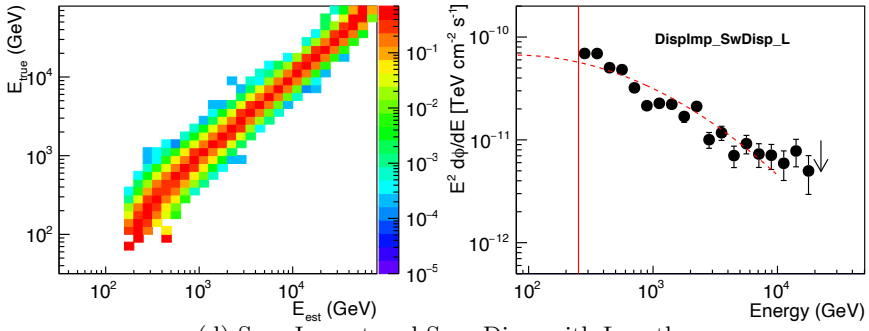
(a) ClasicalImpact and SwapDisp, without Length



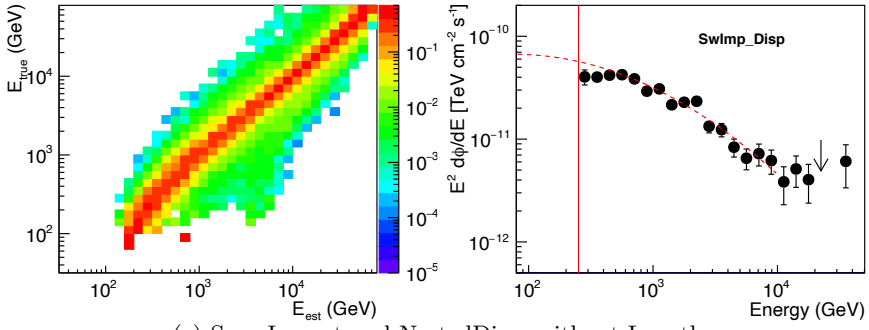
(b) ClasicalImpact and SwapDisp, with Length



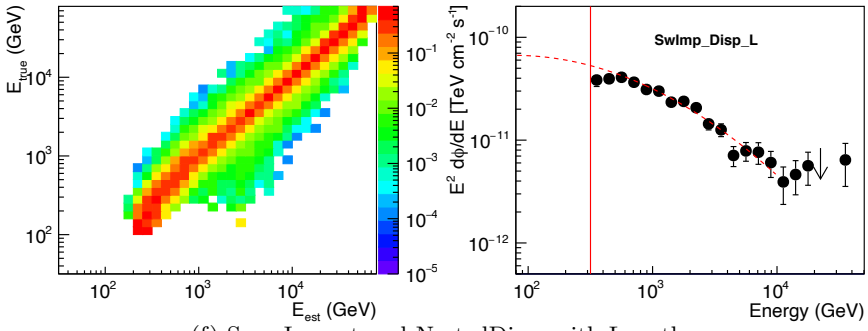
(c) SwapImpact and SwapDisp, without Length



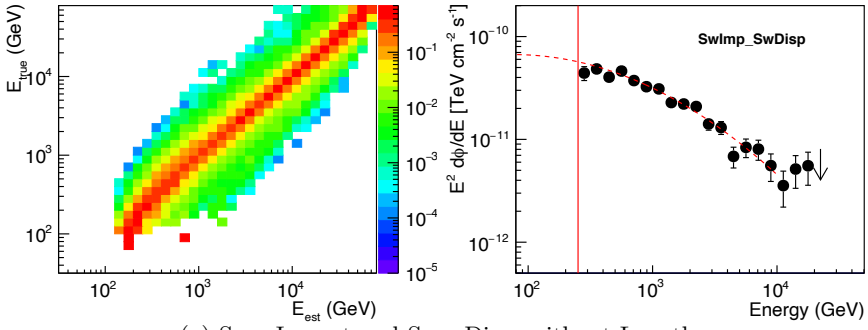
(d) SwapImpact and SwapDisp, with Length



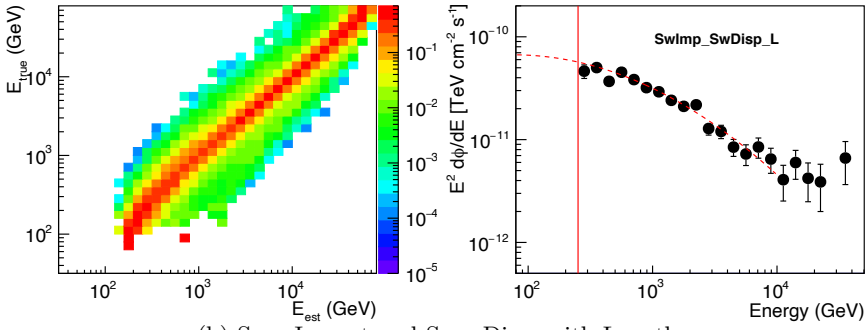
(e) SwapImpact and NestedDisp, without Length



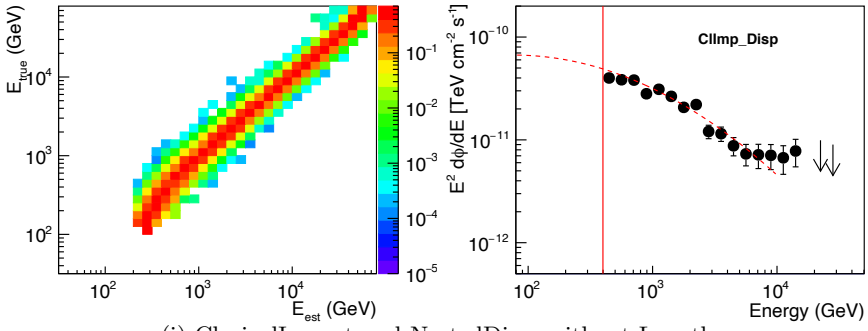
(f) SwapImpact and NestedDisp, with Length



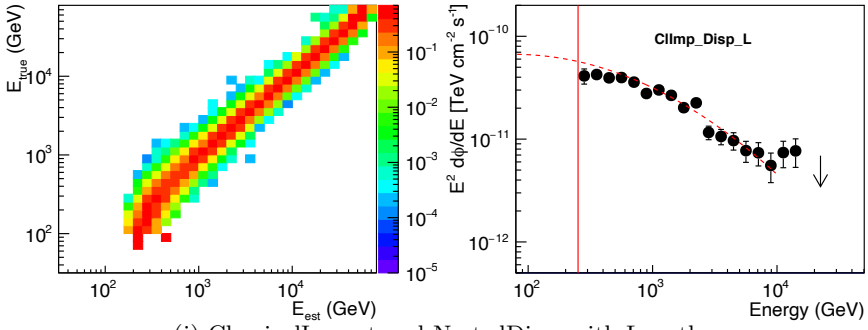
(g) SwapImpact and SwapDisp, without Length



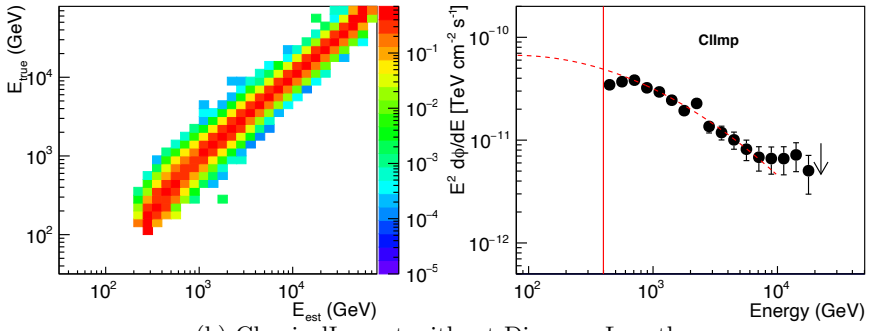
(h) SwapImpact and SwapDisp, with Length



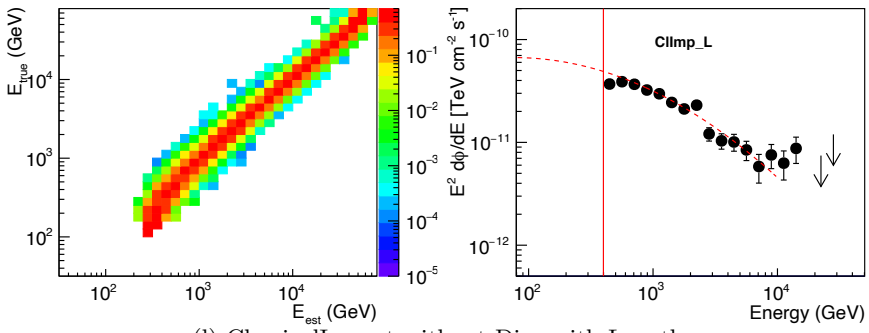
(i) ClasicalImpact and NestedDisp, without Length



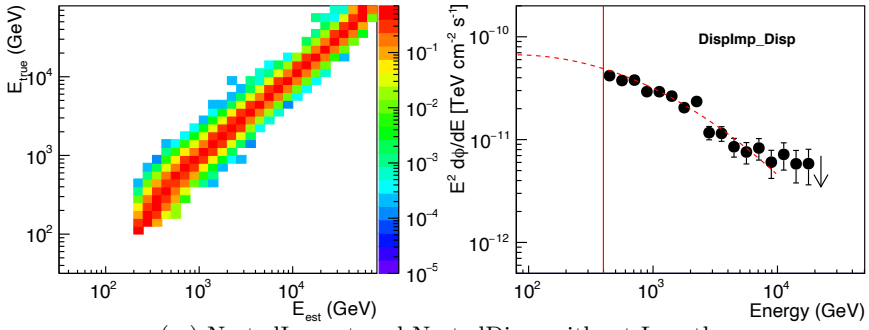
(j) ClassicalImpact and NestedDisp, with Length



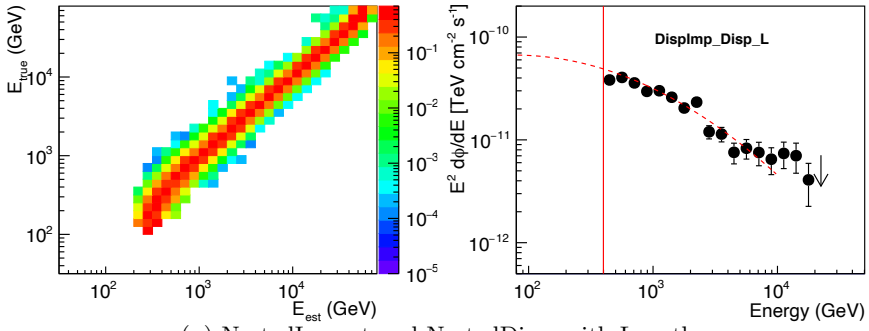
(k) ClassicalImpact without Disp nor Length



(l) ClassicalImpact without Disp with Length



(m) NestedImpact and NestedDisp, without Length



(n) NestedImpact and NestedDisp, with Length

Figure E.11: The migration matrices and Crab Nebula spectra by the RF energy estimators with different input variables

The Crab Nebula spectra obtained with RF-Erec (right) and the migration matrix (left), from the observation data with the Zd 55-65 deg. The red dashed line is the Crab Nebula spectrum presented in past study by MAGIC collaboration [16]. Based on RF-Erec-v2, the input variables are changed. About Impact, the followings are used, "ClImp": ClassicalImpact, "SwImp": SwapImpact (Swap use of DispImpact), "DispImp": DispImpact(Nested use of DispImpact). About Disp, "SwDisp": SwapDisp(Swap use of Disp), "Disp"(Nested use of Disp). And Length is added for "L".

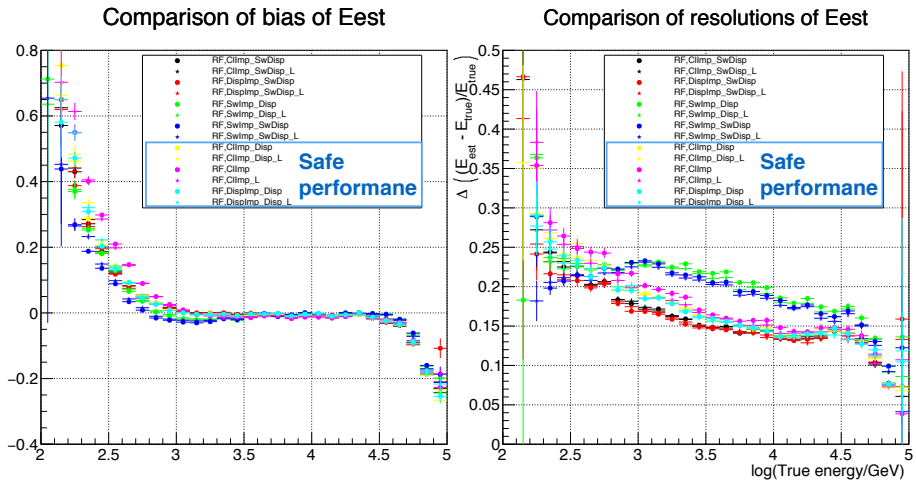


Figure E.12: The performances of the RF energy estimators with different input variables

Based on RF-Erec-v2, the input variables are changed. About Impact, the followings are used, "ClImp": ClassicalImpact, "SwImp": SwapImpact (Swap use of DispImpact), "DispImp": DispImpact(Nested use of DispImpact). About Disp, "SwDisp": SwapDisp(Swap use of Disp), "Disp"(Nested use of Disp). And Length is added for "L". The energy estimators with safe performance, namely without visible systematic effect in the SED, is listed in the legend.

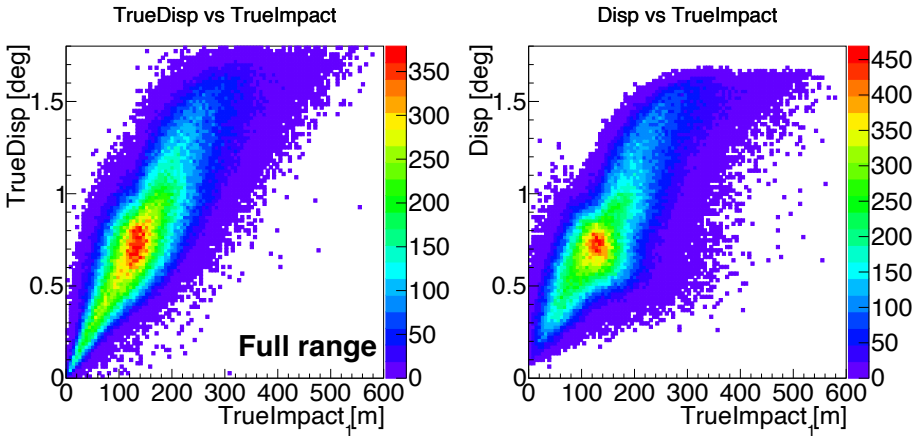


Figure E.13: Disp vs. TrueImpact, for TrueDisp and estimated Disp

Figure E.13 shows such comparison, and it shows the systematic effect of estimated Disp. Compared to the original relation between TrueDisp and TrueImpact, estimated Disp tends to have two sides around the boundary of Impact $\sim$ 130 m with opposite trends. The smaller Impact side tend to have larger estimated Disp than TrueDisp, while larger Impact side tend to be smaller, and both sides become suppressed towards around 0.7, which is the similar to the peak value of TrueDisp for the TrueImpact around 130 m. This Impact is the event with the edge of light pool hitting at the telescope. The smaller Impact indicates the location of telescope to be inner region of the light pool, while the larger Impact is outside. For the inner region of the light pool, as figure 4.6 indicates, the brightness would not be so different and the shift of Disp would not heavily affect the energy estimation. However, the brightness drops as approximately inverse proportional to Impact in outer region, therefore the shift of Disp heavily affect the energy estimation. The smaller Disp indicates the smaller Impact, thus the estimated energy would be lowered. The estimated energy can be lowered only until the lowest energy possible, therefore the drop of estimated energy is less influential in the lowest energy range. Consequently the events are accumulated around the lowest energy and they create the decrease of the excess events at around 1 TeV and increase at the lowest energy.

Appendix F

Application of Neural Network as cross check

Artificial Neural Networks (ANN), also called neural networks or neural nets, are nonlinear computing models vaguely inspired by the biological neural networks that constitute animal brains. Over the last decade, ANN have emerged as the one of most powerful and widely-used supervised learning techniques. For more details, see [29, 84].

Here I discuss the study on the energy reconstruction using ANN. In the beginning phase of my project, I tried the improvement of the energy reconstruction by both RF and ANN. I achieved good performance in the energy reconstruction also with ANN, with the same set of input variables as RF-Erec-v1. It can perform with similar accuracy as RF-Erec-v1. However I decided to give further effort on RF-Erec for improvement, because the performance of RF-Erec is better and there are the more advantages in RF-Erec listed in section 4.2.

In the following, I show the comparison of the performance between ANN and RF-Erec-v1. I also show one of the most important disadvantage in ANN, which is dependency of the performance on filtering the training samples.

Comparison of Neural Network with Random Forest

The basic functionality of ANN was implemented in standard MAGIC analysis software package, as a library based on JETNET [93]. However the performance had not been significantly better than LUTs-Erec. Choosing the exact network architecture for a ANN remains an art that requires extensive numerical experimentation and intuition, and is often times problem-specific. Both the number of hidden layers and the number of neurons in each layer can affect the performance of a neural network. There seems to be no single recipe for the right architecture for a neural net that works best.

In my trial, the better performance was achieved mainly by changing the target value to be in the form of logarithm base 10. Other than this point, the configuration of the ANN energy estimator I applied is as below.

- Input variables
  Input is the same as RF-Erec-v1.

- Cut condition
  The training samples are filtered. Compared to RF-Erec, the filtering is more important factor to tune the performance. However the cut condition I introduce here is not thoroughly optimized. The cut conditions are in the followings[1] ; the number of islands to be 1, the core pixels [2] to be more than 2, Leakage1 to be less than 0.2, Cherenkov Radius to be more than 40 m, Cherenkov Density to be positive, $0 < \theta^2 < 0.1$.

- Activation function
  Sigmoid function is applied for all the nodes.

- Node structure
  The structure of the network has 5 layers. From the input layer, through three hidden layers, to the output layer, the number of nodes are 15, 12, 9, 5, 1.

- Training
  The training is to adjust weights in the nodes so that the predicted value from the input variables matches the target value. The adjustment is done for every event in the train samples after the cut, in which the weights are moved in the direction of the gradient of the cost function. The training should continue until the network stabilization and the number of iterations over whole samples is called epoch. The epoch is 2500 in this study.

The result is compared with the RF-Erec-v1 in figure F.1. The performance is comparable to the RF-Erec-v1, especially in the resolution. Looking in detail, there are the differences in the bias and standard deviation (SD). The bias is smaller (better) in lowest and highest energy side, on the other hand, the standard deviation (SD) is significantly larger (worse). This drawback in SD gives high risk in the analysis for the data with low statistics, therefore RF-Erec is preferred.

Nonetheless, ANN is another independent strategy to energy estimation, and can be used as the opportunity of cross-check.

The need of event filtering

In the training of the ANN, the training samples need to be filtered by proper cut conditions to obtain a good performance of energy estimation. I show this point, and discuss that this is one of the most important reason not to adopt ANN for energy estimation.

Figure F.2 shows the performances of the two ANN energy reconstructions. One is trained with event filtering for the training samples, and the other is done with all the samples. The performance clearly changes with the difference of the event filtering, therefore the event filtering is heavily involved in the performance optimization. On the other hand, RF-Erec does not require optimization via filtering the train samples, and if the same cut is applied to RF-Erec, the performance drops.

[1] For the description of parameters, see section 3.7.

[2] The number of pixels which detected the light content above Q_{core}. The definition is described in section 3.6.

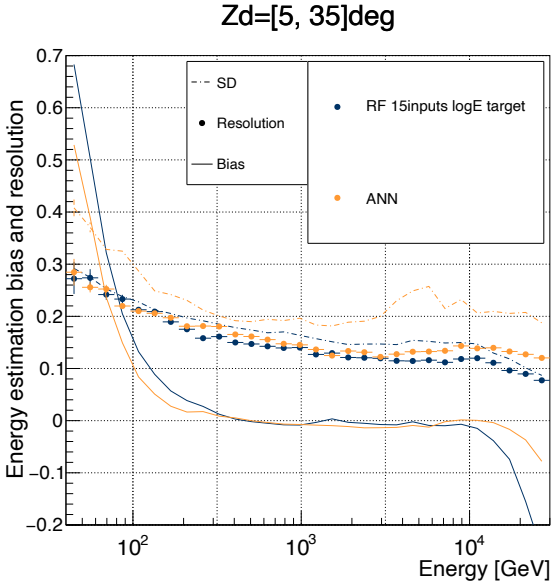


Figure F.1: The performances of the ANN and RF energy reconstructions

The performances of the energy reconstruction by ANN and RF are plotted. The bias (solid line), resolution (the filled circles), and the standard deviation (SD, dashed line) as a function of energy are drawn all together. The definition is explained in section 4.1.3.

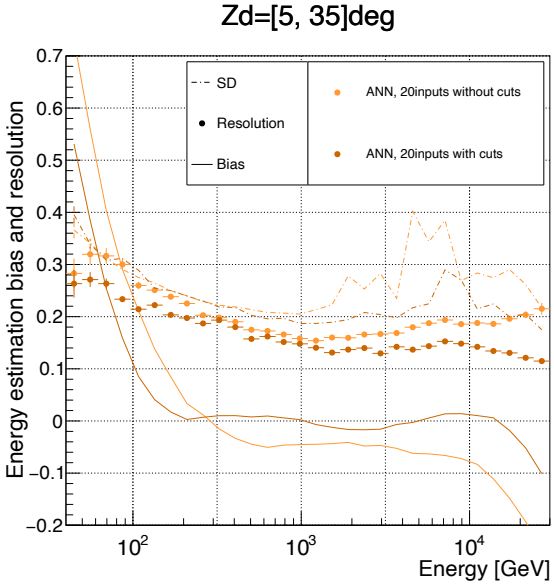


Figure F.2: The performances of the ANN energy reconstructions with and without filtering the training samples

See main text for the cut conditions.

Therefore the training samples need to be filtered properly in the training of the ANN. This means the training process of ANN is sensitive to outlier events and the risk of systematic effect is higher. It is also likely that the optimization to a data set creates a systematic effect to another data set when strict filtering is applied. Therefore it would require more attention on the risk of systematic effects, because the parameter distribution changes for different configurations like Zenith distance, energy range, incoming direction distribution, and so on. Moreover, adjusting the cut condition complicates the optimization of the energy estimator.

Appendix G

Physics related to GRBs

G.1 Prompt emission

Temporal properties (duration and light curve)

The most important temporal properties of the prompt emission are the duration and the light curve. The duration of the prompt emission is associated with the physical origins, and the variability of light curve is the most important clue to discriminate the prompt emission phase from afterglow phase from theoretical point of view.

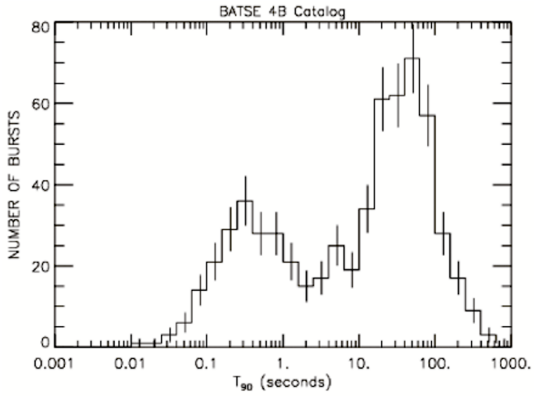


Figure G.1: T_{90} distribution

Duration (T_{90}) distribution of GRBs for the events detected by Burst And Transient Source Experiment (BATSE) onboard *CGRO*. Reproduced from the BATSE GRB 4B catalog[1].

Figure G.1 shows the distribution of the duration of prompt emissions. The duration is characterized with the parameter T_{90}, which is defined as the time interval between the epochs when 5 % and 95 % of the total fluence is registered by the detector. As introduced in section 5.1.1, the duration of the prompt emission distributes in two Gaussian components with a separation line around

[1] https://gammaray.nsstc.nasa.gov/batse/grb/duration/

2 s; a long-duration class with T_{90} peaking at 20–30 s, and a short-duration class with T_{90} peaking at 0.2–0.3 s. It is notable that the relative significance of the two components and the peak duration values are energy and sensitivity dependent. One important property related to the dependence is the spectral hardness of the burst. Statistically, the long-duration class is on average softer than the short-duration class. As a result, the two classes of GRBs are also termed "long/soft" and "short/hard", respectively.

As the examples shown in figure G.2, the light curves of GRB prompt emissions are very irregular. Some are extremely variable, with detectable minimum variability time scale reaching millisecond range, while some others have smooth light curves with relatively simple temporal structures. Some GRBs have distinct emission episodes separated by long gaps in between.

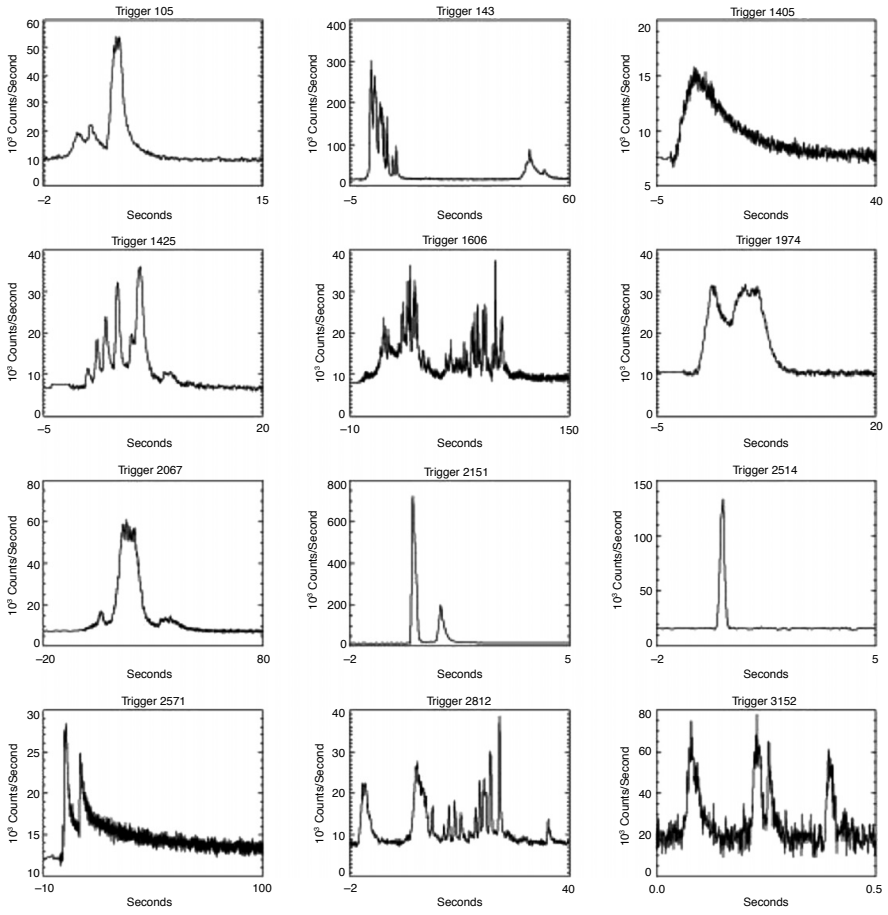


Figure G.2: The examples of prompt emission light curve

The light curves of GRB prompt emissions, detected by BATSE, are drawn. They are labeled by their trigger numbers instead of the dates when they occurred. Figure extracted from [77].

The light curves of GRB prompt emission vary with energy. Figure G.3 shows the light curves of GRB 090902B, observed with Gamma-ray Burst Mon-

itor (GBM) and LAT onboard *Fermi*. In general, pulses tend to be narrower in harder energy bands, and wider in softer energy bands. And arrival time also has a trend with energy. In the energy range below 10 MeV, the arrival time of a pulse in a softer band is typically delayed with respect to the arrival time in a harder band. This trend is seen in Long GRBs in general [120].

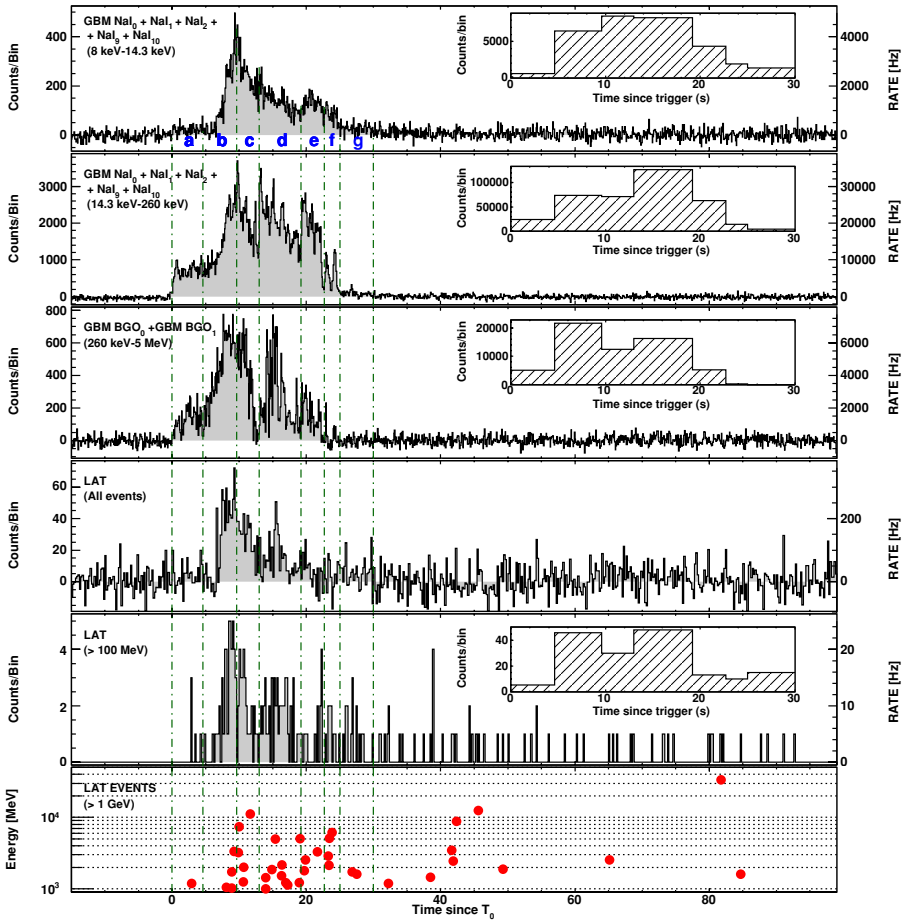


Figure G.3: The examples of prompt emission light curve

Multi-wevelength light curves of GRB 090902B as observed with *Fermi* GBM and LAT. Figure extracted from [3].

Spectrum

The three components in a prompt emission spectrum are; (I) a non-thermal Band component, (II) quasi-thermal component, and (III) another non-thermal component extending to high energies (figure G.4). Usually the component I is the dominant component.

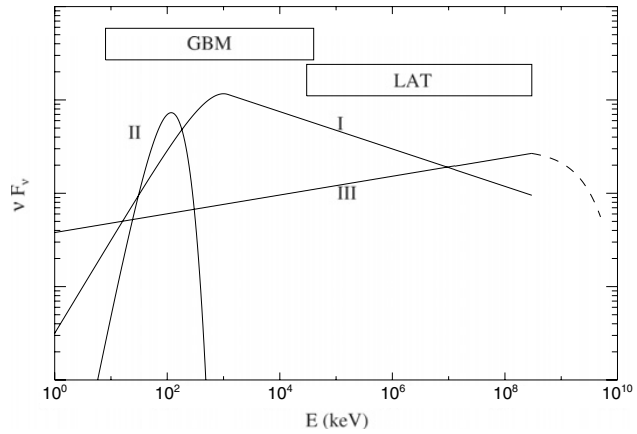


Figure G.4: Prompt emission spectrum with the possible components

The three possible elemental spectral components that shape the observed time-resolved spectra of GRBs. Some components can be suppressed in some GRBs. Figure extracted from [120].

G.2 Afterglow

G.2.1 Light Curve

The X-ray afterglow light curve was summarized from the accumulated samples as a canonical light curve [119], shown in figure G.5. The curve is composed of five possible components, which may or may not exhibit. The components has been discussed as below [26, 77, 117].

- **I. An early time steep decay phase**
  Temporal decay index steeper than -2 and normally -3. This phase is connected to the tail of the prompt emission, and the very rapid decline of emission could be explained only when the jet opening angle is larger than the collimation angle of the emission due to the relativistic effect, Γ^{-1}. The collimation makes the prompt emission from higher latitude reach the observer, however the arrival time will be later because of the longer trajectory, which results in the the tail of temporal decay.

- **II. Shallow decay phase (or plateau phase)**
  The temporal decay of flux is shallow with slope from ~ 0 to ~ -0.7, sometimes even slightly rising early on. Can be incorporated within the external shock model, with the shallow decay phase being due to continuous energy injection into the blast wave. A possible source of energy injection is the power emitted by a spinning-down newly-born magnetar, which explains the relation between the plateau luminosity and the end time of the plateau. Another possibility is the accretion, which could occur if the progenitor star has a core-envelope structure.

- **III. Normal decay phase**
  Temporal decay index ~ -1, which is the typical value predicted in the

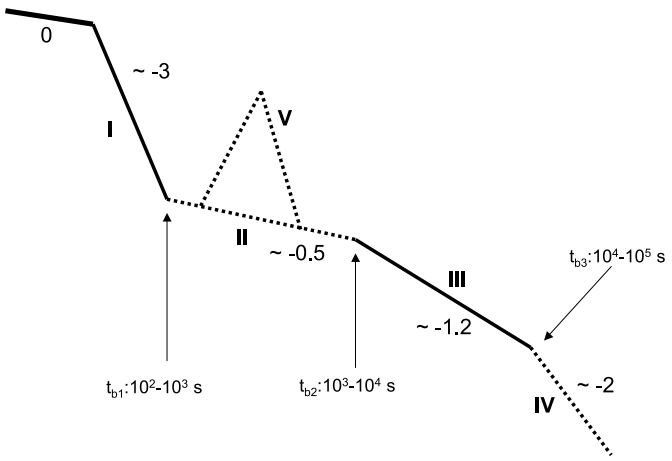


Figure G.5: The canonical X-ray afterglow light curve

The numbers correspond to the 5 distinct components: I. the steep decay phase which is the tail of prompt emission; II. the shallow decay phase (or plateau); III. the normal decay phase; IV. the late steepening phase; V. X-ray flares. The Numerical value provided for each segment of the light curve is the typical decay index for that segment, e.g. the light curve decays as t^{-3} during Phase I. Figure extracted from [119].

standard forward shock model. It is therefore considered as "normal". In theory, the deceleration of the forward shock due to the circumburst medium is the cause of the decay. The actual temporal decay index is dependent on the density profile of circumburst medium and the injection of the high-energy relativistic electrons.

- **IV. Late steep decay phase**
  Temporal decay index t^{-2} or steeper. The decay is expected to follow the normal decay with steeper decay index in the forward shock model. When the jet is decelerated, the collimation angle Γ^{-1} becomes larger. When the angle is no longer small enough, it becomes insufficient to collimate most of the emission towards observer, and the fraction of the emission to reach starts to drop. The start of the steep decay is called jet break.

- **V. X-ray flares**
  The rapid rise and fall, before and after which can be smoothly connected with the same temporal decay index. Therefore it is regarded as flare component superposed on the power-law decay component. One or more X-ray flares can be found in nearly half of GRB X-ray afterglows. Most flares happen early but some flares can be very late, e.g. as late as 10^6 s. These flares share many properties with prompt emission pulses thus are believed to be internal origin.

Figure G.6 shows a set of examples of major light curves with the components listed above. The examples are with no occurrence of the flare component however the light curves with flares comprise of the same types.

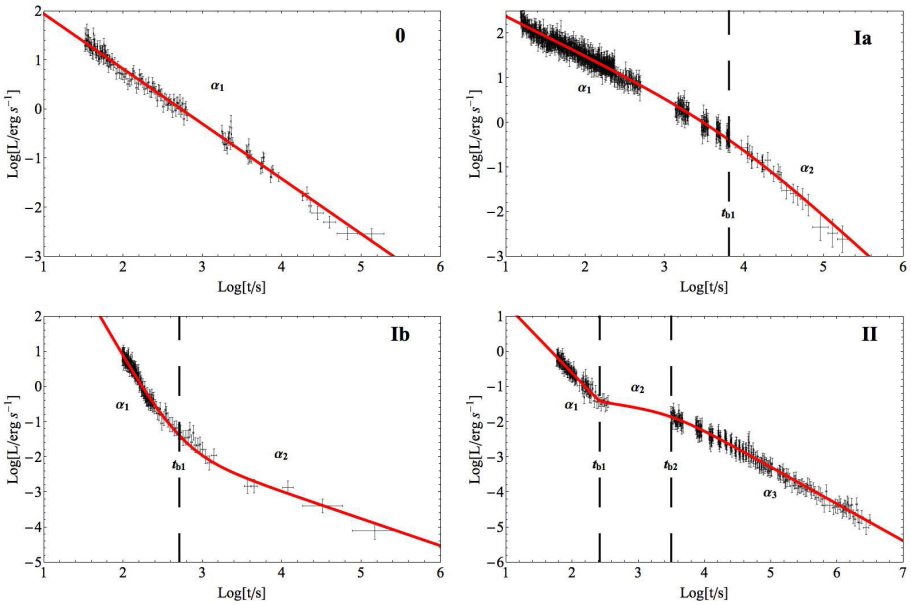


Figure G.6: Examples of X-ray afterglow light curves without flare activity

Examples of the light curves, the luminosity as a function of time in double-logarithmic form, are shown. Upper left, GRB080804, comprises of only normal decay phase (III). Upper right, GRB081222, has shallow decay phase (II) in the beginning and normal decay phase (III) later on. Lower left, GRB090814A, shows early time steep decay phase (I) in the beginning and shallow decay phase (II) later on. Lower right, GRB061021, comprises of early time steep decay phase (I), shallow decay phase (II) and normal decay phase (III). Figure is taken from [26].

G.2.2 Spectrum

Here I summarize the main emission process in afterglow phase to explain the spectrum. The emission is originated from electrons which are swept by the shock wave of the jet. The bulk motion of the shock wave is expressed in bulk Lorentz factor Γ. The electrons in the ambient are swept up by the shock and are accelerated into a power-law distribution. Namely, the energy distribution of the electrons in the co-moving frame of the shock wave is described by the power-law $\frac{dN}{d\gamma_e} \propto \gamma_e^{-p}$, where p is the spectral index and γ_e is the electron Lorentz factor.

On the other hand, due to the enhanced magnetic field in the shock wave, the electrons emit synchrotron radiation and are cooled down. The spectrum of the synchrotron radiation reflects the property of the electrons. In the mean time, the photons emitted are up-scattered by the electrons themselves, and it results in another component in the energy distribution of photons.

In the following I roughly summarize the process above, which is performed in the spectrum modeling. The prime ($'$) refers to quantity measured in the rest frame of the shocked fluid.

Circumburst medium

The density profile of circumburst medium affects the deceleration of the external shock, therefore it is crucial element in modeling the evolution of the emission with time. The profile of the number density of circumburst medium n is formulated as a function of the distance from central engine R,

$$\begin{aligned} n(R) &= n_0 R^{-s} \\ n_0 &= 3 \times 10^{35} A_* \end{aligned} \tag{G.2.1}$$

where A_* is a parameter characterizing the normalization of the density. Two normal profiles are $s = 0$ (homogeneous medium) and $s = 2$ (wind-like medium, typical of an environment shaped by the stellar wind of the progenitor).

Shock wave and particle acceleration

The afterglow is radiated by ensemble of electrons in forward external shock. The radiation can be high energy and non-thermal, because the shock proceeds the very efficient acceleration called first-order Fermi acceleration mechanism. The shock sweeps the electrons in the circumburst medium and generates the power-law energy distribution. Hereafter we discuss such a power-law energy distribution of electrons in the shock wave,

$$\frac{dN_e}{d\gamma_e} \propto \gamma_e^{-p}, \quad \gamma_m < \gamma_e < \gamma_M \tag{G.2.2}$$

where γ_m and γ_M are the minimum and maximum Lorentz factors of the electron energy distribution, respectively.

The minimum Lorentz factor can be estimated from the relation between shocked and unshocked regions. The shock front separates the un-perturbed region (upstream) and the perturbed region (downstream). The unshocked upstream is cold and will be heated by the passage of the shock front. In other words, the number density and internal energy increases, and the effect is enhanced if the shock propagates at relativistic speed. Let us denote the number density in upstream and downstream as n_1 and n_2 (effectively the particles are protons with mass m_p), and the internal energy in upstream and downstream as e_1 and e_2. The shockwave propagates at the bulk Lorentz factor Γ. Using the conservation laws of mass, momentum and energy across the shock front,

$$\begin{aligned} n_2 &= (4\Gamma + 3) n_1 \simeq 4\Gamma n_1 \\ e_2 &= (\Gamma - 1) n_2 m_p c^2 \simeq 4\Gamma^2 n_1 m_p c^2 \end{aligned} \tag{G.2.3}$$

The internal energy e_2 is converted to the electrons' energy and magnetic field energy at certain efficiency. Let ϵ_e the fraction of the shock internal energy that is partitioned to electrons, and ϵ_B the one to magnetic fields. The electrons in the shock can be denoted as

$$\begin{aligned} \int_{\gamma_m} N(\gamma_e)\, d\gamma_e &= n_2 \\ m_e c^2 \int_{\gamma_m} N(\gamma_e) \gamma_e\, d\gamma_e &= \epsilon_e e_2 \end{aligned} \tag{G.2.4}$$

Using the relations above and assuming $p > 2$, γ_m is obtained as

$$\gamma_m = \epsilon_e \frac{p-2}{p-1} \frac{m_p}{m_e} \Gamma \tag{G.2.5}$$

Also the magnetic field B can be obtained

$$B = (32\pi m_p \epsilon_B n_1)^{1/2} \Gamma c \tag{G.2.6}$$

Synchrotron radiation

Synchrotron radiation, electromagnetic radiation of relativistic particles in a magnetic field, is widely believed to power GRB afterglow, and to be the leading candidate radiation mechanism of GRB prompt emission as well. In the following, I will briefly introduce the formation of the observed synchrotron spectrum.

The spectrum of the synchrotron radiation radiated by a single charged particle (The derivation of the equations, see [80, 97]) is

$$\begin{aligned} P(\nu, \gamma) &= \frac{\sqrt{3} q^3 B \sin\alpha}{mc^2} F\left(\frac{\nu}{\nu_{\rm ch}}\right) \\ F(x) &\equiv x \int_x^\infty K_{5/3}(\xi)\, d\xi \end{aligned} \tag{G.2.7}$$

where α is pitch angle (the angle between the direction of magnetic field and that of the charged particle), and γ, m and q are the Lorentz factor, the mass and the charge of the charged particle, respectively. $\nu_{\rm ch}$ is critical frequency as

$$\nu_{\rm ch} = \nu_{\rm ch}(\gamma) = \frac{3}{4\pi} \gamma^2 \frac{eB_\perp}{m_e c} \tag{G.2.8}$$

and $K_{5/3}(\xi)$ is modified Bessel function as

$$K_\iota(x) = \int_0^\infty e^{-x\cosh t} \cosh(\iota t)\, dt. \tag{G.2.9}$$

The spectrum is continuous and has a maximum at $x = 0.29$. The $F(x)$ can be approximated in large or small x, so that

$$F(x) \sim \begin{cases} \frac{4\pi}{\sqrt{3}\Gamma(\frac{1}{3})} \left(\frac{x}{2}\right)^{1/3} & \text{for } x \ll 1 \\ \left(\frac{\pi}{2}\right)^{1/2} e^{-x} x^{1/2} & \text{for } x \gg 1 \end{cases} \tag{G.2.10}$$

The spectral shape is shown in top left of figure G.7. The spectrum has a peak ν_{ch} with rising part with $F(\nu) \propto \nu^{1/3}$ and steep falling part.

The total radiation power is dependent on the energy (represented by the Lorentz factor γ). Integrating eq.(G.2.7) over ν, the total emission power of the particle reads

$$P(\gamma) = 2\sigma_T c \gamma^2 \beta^2 U_B \sin^2\alpha \tag{G.2.11}$$

where U_B is the magnetic field energy density in the emission region[2], and σ_T is Thomson scattering cross section[3]. This radiation power can be interpreted that the synchrotron radiation has the cooling mechanism with the relation $\dot{\gamma} \propto \gamma^2$, and the radiation immediately cools down with the time scale $t_{\rm cool} = 7.8 \times 10^8/(B\gamma)$ s.

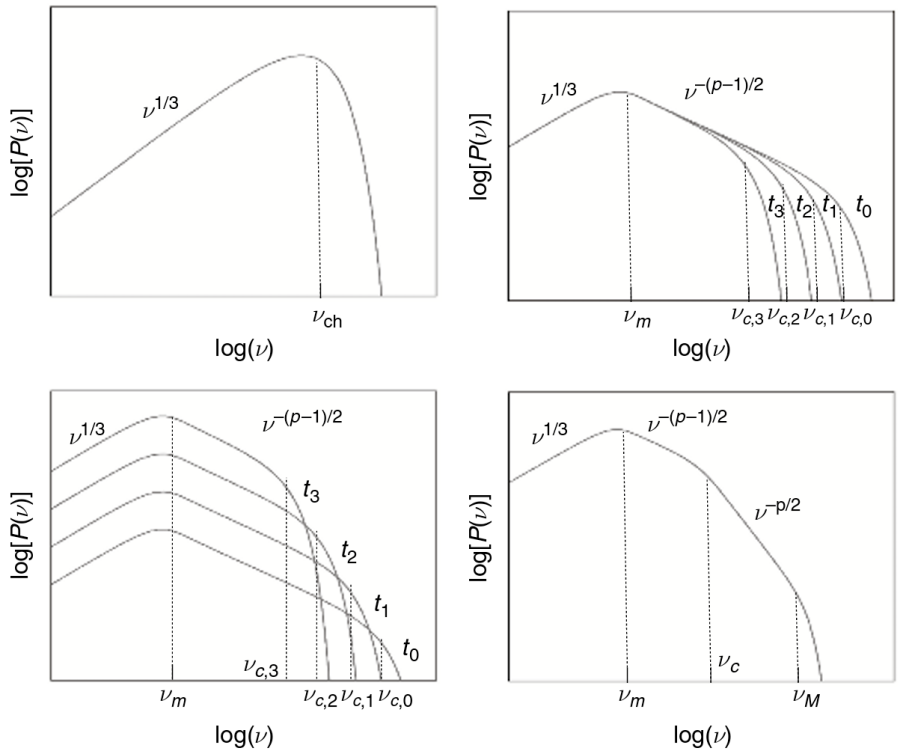


Figure G.7: Synchrotron spectra of electrons

Synchrotron spectra of electrons. Top left: Spectrum for one single electron (or mono-energetic electrons) in a uniform magnetic field. Top right: Spectrum of an ensemble of electrons with a power-law energy distribution with an impulsive injection. A power-law segment is formed, with the high-energy end decreasing with time due to synchrotron cooling. Lower left: When continuous injection of fresh particles is considered, the observed spectrum is a superposition of many spectra produced by electrons with different "ages", forming a new segment. Notice that, for illustration purposes, the injection rate of electrons is adopted to decrease with time. However, the conclusion applies to more general cases (e.g. the injection rate of electrons remains the same or increases with time). Lower right: Putting all these effects together, the spectrum can be described by a three-segment broken power law. Figure extracted from [77].

[2] $U_B = B^2/8\pi$

[3] $\sigma_T = \frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2$.

Synchrotron radiation from ensemble of electrons

The afterglow is radiated by an ensemble of electrons in forward external shock which is distributed as eq.(G.2.2). The total spectrum of synchrotron radiation is superposition of the emission from each particle. Since the radiation immediately cools down the distribution, the spectrum distortion needs to be taken into account.

Top right of figure G.7 is the time development of the synchrotron radiation spectrum from an impulsive injection of the electrons with the energy distribution. The peak is formed at ν_m, correspondent to the minimum Lorentz factor of the injection γ_m, and the slope with the power-law index $-(p-1)/2$ is formed in the right of the peak. The cutoff occurs at ν_M, correspondent to γ_M at the moment of injection, however it will become $\nu_{c,i}$, the critical frequency of the highest Lorentz factor of the distribution at t_i $(i = 0, 1, 2, 3)$. This is due to the faster cooling for higher energy electron.

If the injection is continuous, the observed spectrum is stable but there becomes a spectral break between ν_m and ν_M, shown in the bottom right of figure G.7. The frequency is denoted as ν_c, and the correspondent Lorentz factor γ_c is called cooling energy or cooling Lorentz factor. The power-law index at $\nu_c < \nu < \nu_M$ is $-p/2$.

Synchrotron radiation with fast/slow cooling

If the cooling is fast, the entire spectrum will be affected because of very low γ_c with $\gamma_c < \gamma_m$. In this case, there are two energy range differently affected, which are $\gamma_e > \gamma_m > \gamma_c$ and $\gamma_m > \gamma_e > \gamma_c$.

To examine them, one can consider a continuity equation of electrons in energy space. Associated to the equation $\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) = \sigma$,

$$\frac{\partial}{\partial t}\frac{dN_e}{d\gamma_e} + \frac{\partial}{\partial \gamma_e}\left[\dot{\gamma}_e \frac{dN_e}{dt}\right] = S(\gamma_e) \tag{G.2.12}$$

where $\frac{dN_e}{d\gamma_e}$ is the spectrum of the electron, $\frac{\partial}{\partial \gamma_e}\left[\dot{\gamma}_e \frac{dN_e}{dt}\right]$ is the flow of the electrons in the energy space, and $S(\gamma_e)$ is the source term. Our interest is to see the stable spectrum case, namely $\frac{\partial}{\partial t}\frac{dN_e}{d\gamma_e} = 0$.

In the slow cooling case with $\gamma_e > \gamma_c > \gamma_m$ or the fast cooling case with $\gamma_e > \gamma_m > \gamma_c$, eq.(G.2.12) is

$$\frac{\partial}{\partial \gamma_e}\left[\dot{\gamma}_e \frac{dN_e}{dt}\right] = C_1 \gamma_e^{-p} \tag{G.2.13}$$

The cooling mechanism of synchrotron radiation has the relation $\dot{\gamma} \propto \gamma^2$. The electron spectrum, and accordingly the synchrotron radiation spectrum, are modified in the following.

$$\frac{dN_e}{dt} \propto \gamma_e^{-(p+1)}, \text{ thus } P(\nu) \propto \nu^{-p/2} \tag{G.2.14}$$

In the energy range $\gamma_m > \gamma_e > \gamma_c$ in the fast cooling case, eq.(G.2.12) is

$$\frac{\partial}{\partial \gamma_e}\left[\dot{\gamma}_e \frac{dN_e}{dt}\right] = 0 \tag{G.2.15}$$

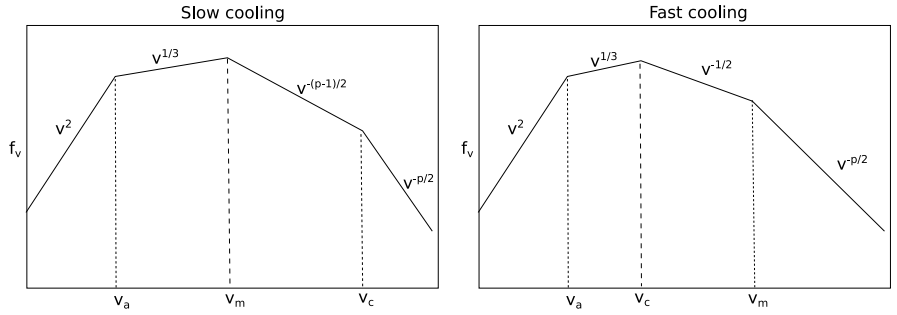


Figure G.8: Synchrotron spectra in fast and slow cooling

In the slow cooling case (left), cooling energy ν_c is above the minimum injection energy ν_m. However, the fast cooling case (right) cools down to below the injection energy, and deforms the spectrum. ν_a is called synchrotron self absorption frequency, below which the emergent specific flux cannot exceed the black-body flux corresponding to the appropriate electron temperature. Figure extracted from [77].

So that

$$\frac{dN_e}{dt} \propto \gamma_e^{-2}, \text{ thus } P(\nu) \propto \nu^{-1/2} \tag{G.2.16}$$

The spectra in fast and slow cooling cases are shown in figure G.8.

The peak of synchrotron emission at the frame of an observer can be denoted by

$$\begin{aligned} \varepsilon_{\text{peak}}^{\text{Syn}} &= \frac{3h}{4\pi}\Gamma\gamma_{e,\text{peak}}^2\frac{eB'}{m_e c} \\ &= 35\left(\frac{\Gamma}{20}\right)\left(\frac{B'}{10\text{G}}\right)\left(\frac{\gamma_{e,\text{peak}}}{10^5}\right)^2 \text{keV} \end{aligned} \tag{G.2.17}$$

where the Doppler boosting effect by the bulk Lorentz factor Γ is taken into account.

Maximum synchrotron energy

The cooling time scale by synchrotron radiation is shorter for higher energy of an electron, however it takes more time to accelerate it to higher energy, thus there is a maximum energy for electrons. There are several ways to obtain the maximum energy, but the simplest way is as follows.

The minimum time required for acceleration of an electron while crossing a shock front is of the order of the Larmor time

$$t'_L = m_e c\gamma_e/(eB') \tag{G.2.18}$$

In this time scale, the first order Fermi process takes place and electrons gain energy by a factor $\sim$2 each time they are scattered from one side to the other of a relativistic shock front. Therefore the particle should not lose more than half its energy to synchrotron radiation in time t'_L.

Equating that the energy loss of an electron by synchrotron radiation [4] for the time t'_L is smaller than half of the electron's energy,

$$\frac{4}{3}c\sigma_T\beta^2\gamma^2\frac{B'^2}{8\pi}t'_L < \frac{\gamma_e m_e c^2}{2} \tag{G.2.19}$$

Therefore, the two equations above yields [5],

$$h\nu_{sync} = h\frac{3}{4\pi}\gamma_e^2\frac{eB'}{m_e c} < \frac{27}{16}m_e c^2\frac{\hbar c}{e^2} \sim 118\,\mathrm{MeV} \tag{G.2.20}$$

The observed spectrum of afterglow synchrotron emission is then expected to display a cutoff below the energy $\varepsilon_{\max}^{\mathrm{Syn}} \sim 100\,\mathrm{MeV} \times \Gamma/(1+z)$.

Synchrotron Self-Compton(SSC) emission

A high energy particle can transfer its energy to lower energy photon through collision process, called IC scattering process. Besides synchrotron radiation, IC process by high energy electron is another important radiation mechanism in high energy phenomena to generate high energy photons. SSC is the IC process in a distribution of high energy electron, where high energy electrons up-scatter the photons emitted by the electrons in the same distribution via synchrotron emission.

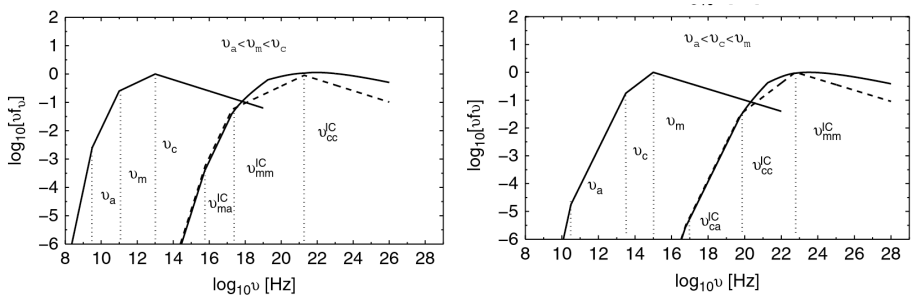


Figure G.9: Synchrotron+SSC spectra in fast and slow cooling

Synchrotron and SSC spectra of electrons with slow cooling regime (left) and fast cooling regime(right). Figure extracted from [54].

From a high energy electron with Lorentz factor γ_e, a typical photon gains energy by a factor of γ_e^2 and the collision cross section is almost constant by Thomson cross section σ_T. Thus the SSC spectrum reproduces almost the same shape as the spectrum of the seed photon of the synchrotron radiation (figure G.9). However, it is valid only when the photon energy is sufficiently low (Thomson regime) because the photon energy after collision cannot exceed $\gamma_e m_e c^2$ due to energy conservation law. The energy regime with the suppressed

[4] For isotropic distribution of particles, the pitch angle is averaged to be $\langle \sin^2\alpha \rangle = \frac{2}{3}$. Therefore the single particle synchrotron power eq.(G.2.11), averaged for the pitch angle, is

$$\langle P(\gamma) \rangle = \frac{4}{3}c\sigma_T\beta^2\gamma^2 U_B$$

[5] The calculation adopted the fine structure constant, $\alpha_f = \frac{e^2}{\hbar c} = 1/137$, and $\beta \sim 1$.

cross section due to high incident energy of the photon is called Klein-Nishina regime [6].

The SSC peak is

$$\begin{cases} \nu_{\rm peak}^{\rm SSC} \sim 2\gamma_e^2 \nu_{\rm peak}^{\rm Syn} & \text{Thomson regime} \\ \nu_{\rm peak}^{\rm SSC} \sim 2\gamma_e \Gamma m_e c^2 & \text{Klein-Nishina regime} \end{cases} \tag{G.2.21}$$

G.3 Extragalactic Background Light (EBL) absorption

Extragalactic Background Light (EBL) is a cosmic diffuse radiation field. It is mainly composed of the ultraviolet, optical and near-infrared light emitted by stars through the history of the universe. It is the brightest background photon field after the Cosmic Microwave Background (CMB) radiation. It permeates the intergalactic space, and VHE γ-rays can be effectively absorbed during their propagation via photon–photon pair-production interactions with low-energy photons of the EBL.

In environments with very high radiation energy density, photons interact and produce a pair of electron and positron.

$$\gamma + \gamma \rightarrow e^+ + e^- \tag{G.3.1}$$

In the propagation of the VHE γ-rays from extragalactic sources to Earth, the γ-rays traverse in the radiation field of EBL. Therefore they can interact with EBL photons if there is sufficient energy in the center of mass frame to create an electron-positron pair, namely twice the electron mass, $2m_e$.

$$\sqrt{2E_\gamma E_{EBL}(1-\cos\theta)} \geq 2m_e c^2 \tag{G.3.2}$$

where E_γ is the energy of γ-ray photon, E_{EBL} is the energy of EBL photon, and θ is the angle of interaction between the two photons to collide.

The cross section for pair production is

$$\sigma(\beta^*) = \frac{3\sigma_T}{16}(1-\beta^{*2})\left[2\beta^*(\beta^{*2}-2) + (3-\beta^{*4})\ln\left(\frac{1+\beta^*}{1-\beta^*}\right)\right] \tag{G.3.3}$$

where $\beta^* = v^*/c$ with v^* being the velocity of electron (positron) in center of momentum frame. The β^* can be expressed in terms of the energies of the photons to collide E_1 and E_2, and the collision angle θ. Introducing total energy of electron(positron) $E_e^* = \gamma^* m_e c^2 = m_e c^2 (1-\beta^{*2})^{-1/2}$,

$$\begin{aligned} E_1 E_2 (1-\cos\theta) &= 2\left(\frac{1}{1-\beta^{*2}}\right) m_e^2 c^4 \\ \beta^* &= \sqrt{1 - \frac{2m_e^2 c^4}{E_1 E_2 (1-\cos\theta)}} \end{aligned} \tag{G.3.4}$$

[6] Thomson scattering occurs for a photon with a incident energy E_i' in the comoving frame of the electron, when $E_i' \ll m_e c^2$. Namely, in lab frame, $E_i \ll \gamma m_e c^2$

This yields the maximum of cross section,

$$\sigma_{max} \sim 0.25\sigma_T \qquad \text{at } E_1 E_2 \sim 4m_e^2 c^4 \tag{G.3.5}$$

This tells that a 1 TeV photon has the largest cross section with a $\sim$ 1eV photon. As EBL contains large amount of optical radiation, VHE γ-ray is the waveband to suffer from strong absorption. Therefore it is important for the analysis of intrinsic spectrum to estimate the absorption.

The optical depth relates the total distance a γ-ray travels to its mean free path for interaction τ as a full integral over the distance. For γ-rays with energy of E_γ coming from a source at a red shift of z_0, τ is

$$\tau(E_\gamma, z_0) = \frac{1}{2}\int_0^{z_0} dz \frac{dl}{dz} \int_{-1}^{1} d(\cos\theta)(1-\cos\theta) \times \int_{E_{min}}^{\infty} d(E_{EBL}) n(E_{EBL}, z)\sigma(E_\gamma(1+z), E_{EBL}, \theta) \tag{G.3.6}$$

where $\dfrac{dl}{dz}$ is the cosmological element

$$\frac{dl}{dz} = \frac{1}{(1+z)H_0} \frac{1}{\Omega_m(1+z)^3 + \Omega_\Lambda} \tag{G.3.7}$$

There are some models of the EBL available, and the attenuation and in this is correspondingly different. This study adopts the model of Domínguez et al. [44]. Figure G.10 shows the optical depth for the γ-rays at the distance of GRB 190114C. The γ-rays at TeV energies suffer from severe attenuation by a factor of about 300.

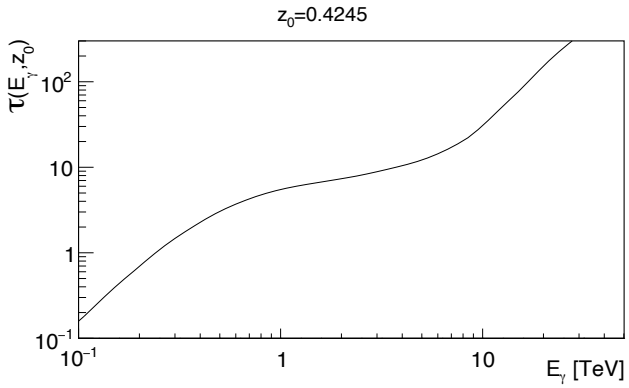


Figure G.10: The optical depth vs the energy of γ-rays for a source at z =0.4245

The optical depth of γ-rays from the source at the distance of $z_0 = 0.4245$, as a function of the γ-ray's energy. The ,calculated with the nominal EBL scale model of Domínguez et al. [44].

Appendix H

Methods used to evaluate the various unfolding strategies

In this chapter, I discuss the methods to evaluate the unfolding strategies. As the basis of unfolding is introduced in section 3.11.2, there are five regularization methods based on different assumptions on the unfolded spectrum. Moreover, the process does not automatically lead to a stable and reasonable solution. The regularization strength, namely the weight ω or the iteration number i, needs to be searched from a wide range, for a reasonable agreement between the assumption and the consistency with the measured values. There are various criteria for choosing the best regularization strength proposed, however none of them is optimal for all cases. Therefore unfolding is a complicated process, and its result needs to be investigated carefully.

Here I introduce the criterion of the search for the best regularization strength and discuss how to evaluate the result using control plots. I also describe the discussion on the control plots needed in section 4.6.4 and section 5.4.2.

H.1 The criteria for the best regularization strength

There are 5 options implemented in the standard analysis procedure, for the search of the best regularization strength. For all the regularization methods, the range of the strength of regularization, ω or i, is set to be from 10^{-5} to 10^{10}. Among the solutions in this range, the best strength of regularization is chosen, judging from χ_0^2 or the noise level. The below are the 5 criterion.

Focusing on χ_0^2

- Least χ_0^2
  The least χ_0^2 is searched as the solution. This means that the regularization is practically ignored and by setting the weight ω in eq.(3.11.5) to be maximum.

- χ_0^2 closest to the number of measured points

Assuming that the number of events measured in each energy bin follows Poisson distribution independently, the χ_0^2 should be around the number of bins.

- χ_0^2 closest to the rank of G
  The χ_0^2 should actually take into account the number of degrees of freedom in the true energy bins. It would be estimated as around the rank of G for the migration matrix.

Focusing on the noise level (covariance matrix)

The noise level, which is the net error of a distribution, should be kept in reasonable range. The noise level in the original measured distribution $\mathbf{y}$ can be regarded as the trace of the covariance matrix K introduced in eq.(3.11.4),

$$Tr(K) = \sum_{i=1}^{n_y} \sigma_{y_i}^2 = \sum_{i=1}^{n_y} y_i \tag{H.1.1}$$

On the other hand, the noise level of unfolded distribution $\hat{\mathbf{s}}$ is obtained as the trace of the covariance matrix T of $\hat{\mathbf{s}}$. When $\hat{\mathbf{s}}$ is obtained using approximate inverse matrix of the migration matrix M as $\hat{\mathbf{s}} = D\mathbf{y}$, the covariance matrix T is

$$T = D \cdot K \cdot D^T \tag{H.1.2}$$

Using these values, there are two following strategies implemented.

- Comparable noise level after unfolding
  The distribution after unfolding should not be too different. This can be interpreted as the noise levels to be similar between before and after unfolding. Therefore the criterion is

$$Tr(T)/Tr(K) = 1 \tag{H.1.3}$$

- Strongest noise increase in unfolded distribution
  The noise level after unfolding is dependent on the strength of regularization. Since regularization is introduced to reduce noise, the noise level increases for weaker regularization, namely the larger ω or i. And it can be a strategy to choose the point of highest gradient of noise level for increasing ω or i.

H.2 Evaluation of the unfolding results

In this section I discuss how to investigate the results from unfolding in the MAGIC standard analysis, which adopts five regularization methods with five criteria of choosing the strength of regularization, introduced in section 3.11.2 and section H.1 respectively.

There is no promising way to obtain the reasonable unfolding result. Instead, the results from those different regularization methods with different criteria should be compared to choose the most reasonable ones. The unfolded spectral data points should be reasonably regularized and must explain the measured number of excess events through migration matrix. To this end, the unfolding results in this study are evaluated for the following points.

Back-folded distribution

The unfolded distribution must be consistent with the measured distribution. When the unfolded distribution is folded again with the migration matrix (back folded), the distribution must agree with the measured distribution (the distribution before the unfolding). This can be examined by the plot of measured distribution together with backfolded distribution overlaid.

χ_0^2 with respect to the number of bins

The overall consistency can be represented by χ_0^2 in eq.(3.11.5), which is the discrepancy between the backfolded distribution and the measured distribution. The migration can be regarded as statistical process, and χ_0^2 should be within fluctuation. As a reasonable value of the fluctuation, I refer to the number of estimated energy bins used for the migration and the rank of G [1].

The noise level in the unfolded distribution

Noise increase, $Tr(T)/Tr(K)$ should be close to 1. If it is too large, the noise is too large and it results in the error matrix with too large uncertainties in the correlation bins.

The strength of regularization

The regularization should not be too strong or too weak.

Agreement among the best-regularization-search criteria

In each regularization strategy, it is needed to choose the best result from the five criteria for searching the best regularization strength. However, three of the criteria are relatively reasonable in general; χ_0^2 closest to the number of measured points, χ_0^2 closest to the rank of G, and the comparable noise levels after unfolding. The results given by them should be robust and consistent to each other, because all of these are the requirement for the reasonable result. On the other hand, the Least χ_0^2 is actually the solution when the regularization is almost neglected, and Biggest noise increase does not guarantee the noise level in the unfolded distribution to be similar to that in the measured distribution.

Therefore, if the results by the three criteria are similar, the unfolding can be regarded to be robust.

Agreement among the regularization methods

The different regularization methods are based on the different assumption on the unfolded spectrum, and some of the results can fail if the assumption is not proper. However the assumptions are reasonable in most cases, thus a good spectral analysis requires the agreement among the different methods.

[1] In the point of view of unfolding process, it is possible to compute the effective number of degrees of freedom, which is the number of bins neglected in solving the unfolding problem. In the unfolding for MAGIC standard analysis, it is computed only in some of the regularization methods.

H.3 Control plots in the performance evaluation

To obtain the result of unfolding, shown in figure 4.44, the estimated energy range and true energy range are selected as shown in figure 4.41. Figure H.1 is the control plots for the results, in which the best regularization strength is determined by the point of comparable noise level.

Tikhonov

While RF-Erec gives consistent results among the different criteria with reasonable χ^2_0 and with reasonable power-fit results, the unfolding with LUTs-Erec results in two tendencies of unfolded spectra, both of which are deviated from the original spectrum.

In noise-based solutions, the spectrum looks good power law, and PL fit is fine($\sim$ -4.96 $\pm$ 0.03), however there is an isolated point at around 10 TeV after two bins with no excess. On the other hand, in the spectra by the χ^2_0 based solutions, there is sudden drop of flux after 1 TeV, which read as a cutoff. These problems are not visible in the other control plots. In all the criteria, the χ^2_0 are mostly good, the smoothness is not significantly bad.

In RF, the data points are consistent among all the weight choice, and the PL fits show the same slope in all criteria for the best regularization strength.

Schmelling

In both RF-Erec and LUTs-Erec, the unfolded spectra are mostly consistent with the original spectrum. The power-law fit to the spectral points gives the same index as original (at most 0.05 pm 0.03 difference).

There is a difference between RF-Erec and LUTs-Erec in the control plot. In LUT, In the middle of the weight range (around 10^3), the Gauss-Newton iteration did not converge for some weights.

And there is a problem in mitigating χ^2_0 and noise level, for both RF-Erec and LUTs-Erec. For the solutions with the criteria for best regularization strength related to χ^2_0, the noise results in bad values by factor of several 10 s. When the comparable noise is chosen for the best regularization strength selection, the results obtain bad χ^2_0 (LUTs-Erec: 100, RF-Erec: 220). In this criterion, the PL fit to the points is also worse, especially in LUTs-Erec (LUTs-Erec: -4.74 $\pm$ 0.02, RF-Erec: -4.90 $\pm$ 0.02).

The same problem happens also to Schmelling(minuit). Schmelling and Schmelling(minuit) may not be proper for a very steep spectrum. They impose the regularization to control the cross entropy, in which the prior distribution is set to be the measured distribution. However, a measured distribution should be heavily distorted from the true distribution because of very steep spectrum. This would also cause the problem that there is no point to satisfy both χ^2_0 and noise level.

Schmelling(m)

The results are in a similar tendency as Schmelling but are worse, especially in LUTs-Erec.

In LUTs, the unfolded data points are deviating from the original spectrum above 1 TeV. The tendency is dependent on the criterion of the best regularization strength. The point around 10 TeV becomes isolated in most of the cases, and in case of comparable noise level, the points show upturn.

The control plots indicate that it is more difficult to mitigate between χ_0^2 and the noise level. It is impossible to obtain both low χ_0^2 and low noise level together.

Bertero

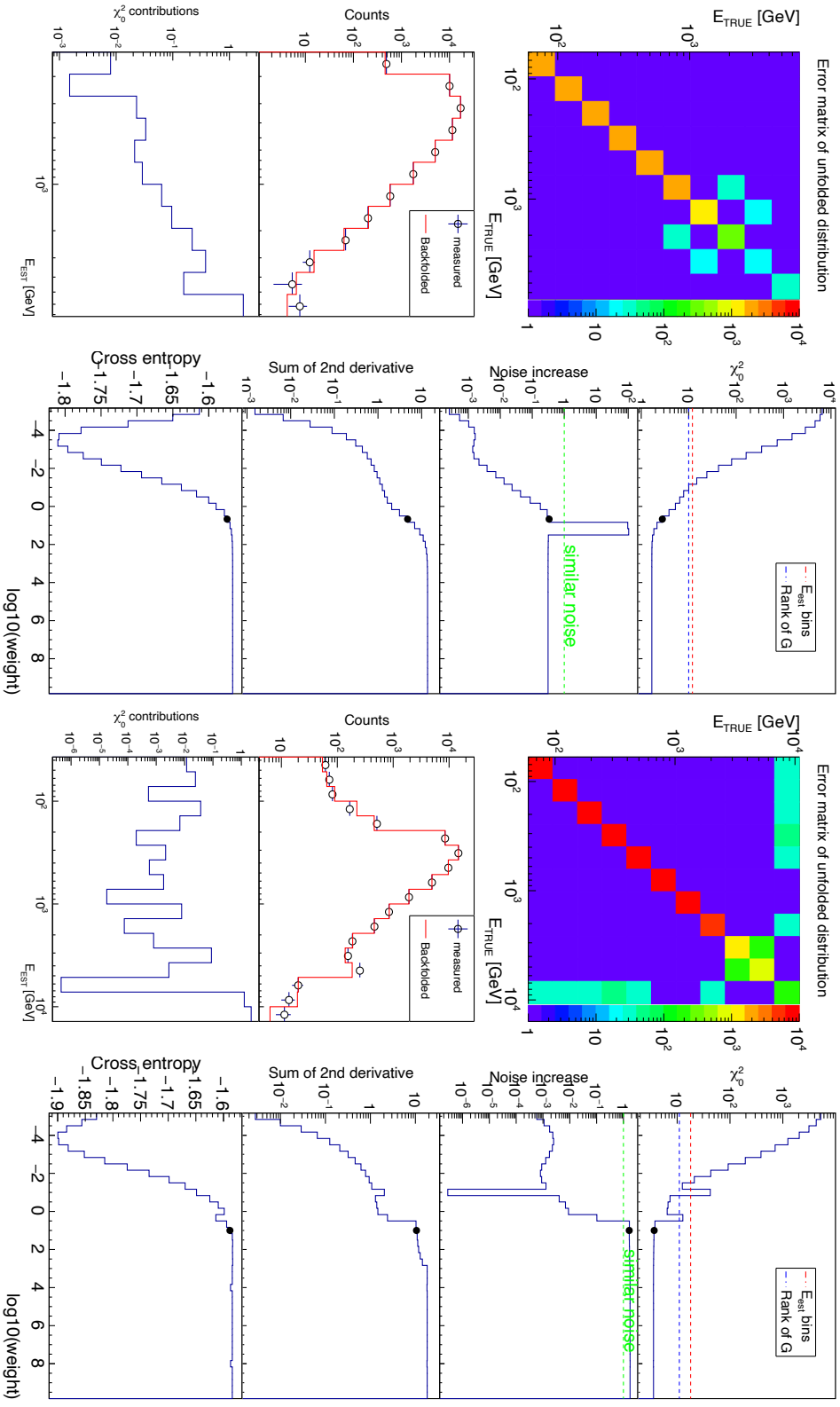
In RF-Erec, all the unfolded spectra are very similar to each other and consistent with the true spectrum. The power-law fits to them result in the same indices and acceptable χ_0^2 and noise levels.

However in LUTs-Erec, the unfolded spectra are totally different for different ways of regularization strength selections, all of which are deviating from the true spectrum except for the choice of comparable noise. Loss of the intermediate points (χ_0^2 = number of measured points), obvious upturn (χ_0^2 = rank of G), and very different PL indices (-3 to -4.71, -4.89) as well as very different chi^2 of the PL fit (1.48, 58,82,2365) among the ways of weight selections. The selection of the best weight does not work properly. The control plots are not stable either. The χ_0^2 as a function of iteration number (the regularization strength) is always over 4000 in unfolding with the criterion of the regularization strength of χ_0^2 = rank G, which means the solution is not reasonable at all.

Bertero(W)

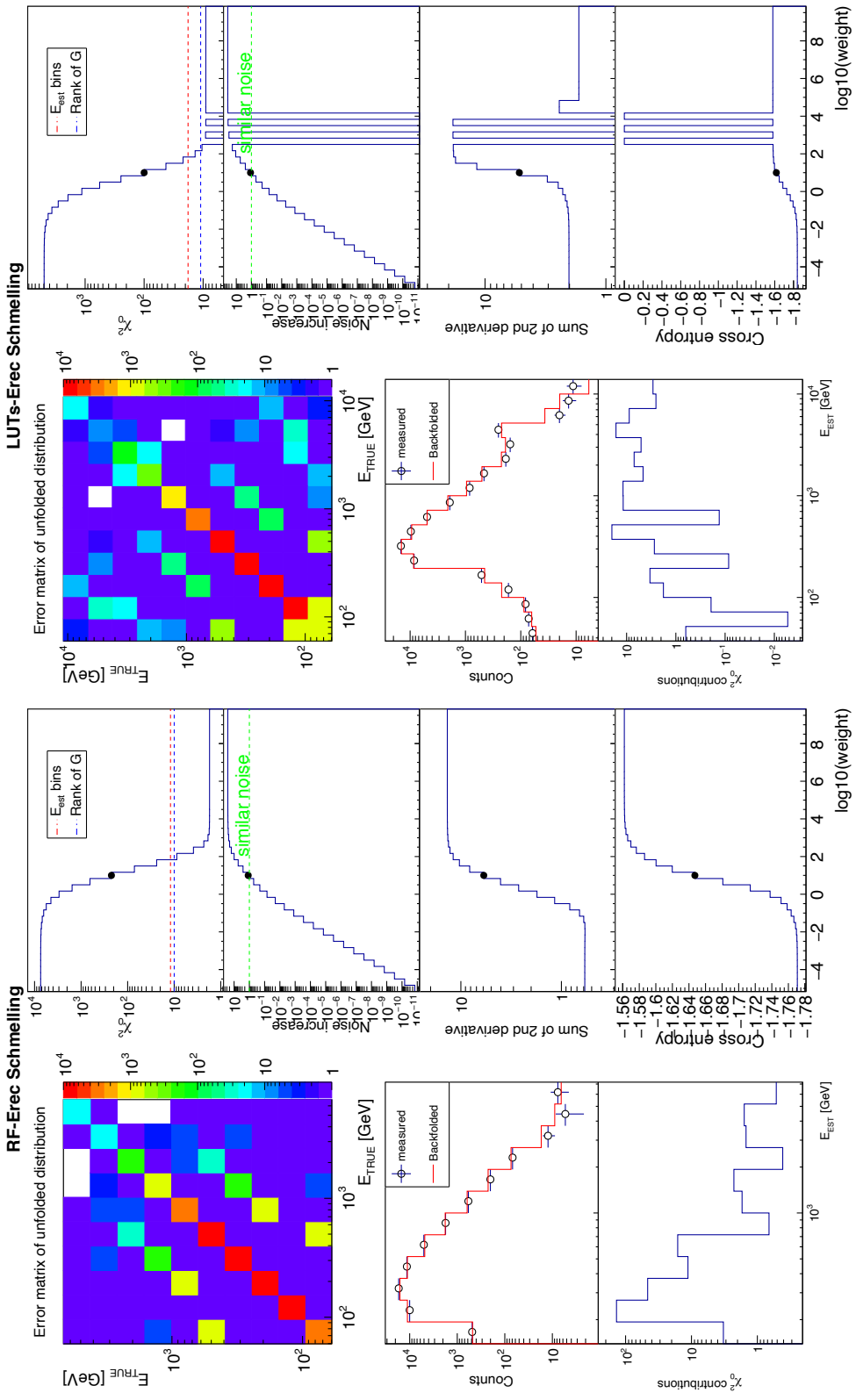
In RF-Erec, the unfolded spectra are similar to those by Bertero method, consistent with the true spectrum for most of the criteria of choosing regularization strength, with an exception that the criterion of comparable noise choice gives harder spectrum (by power -law fit, index is -4.79 $\pm$ 0.02 and $\chi^2/NDF = 159/7$). Looking the control plot, this is due to very fast change of χ_0^2 in regularization strength, while the noise slowly develops. Therefore in this case χ_0^2-based criteria are better.

In LUTs, the results are better than Bertero method, but are still worse than those in RF-Erec. The difference of the unfolded spectra among the different criteria remains. The deviation tendencies are the upturn and isolated high energy point. Consequently the power-law fits give hard indices and bad probabilities.



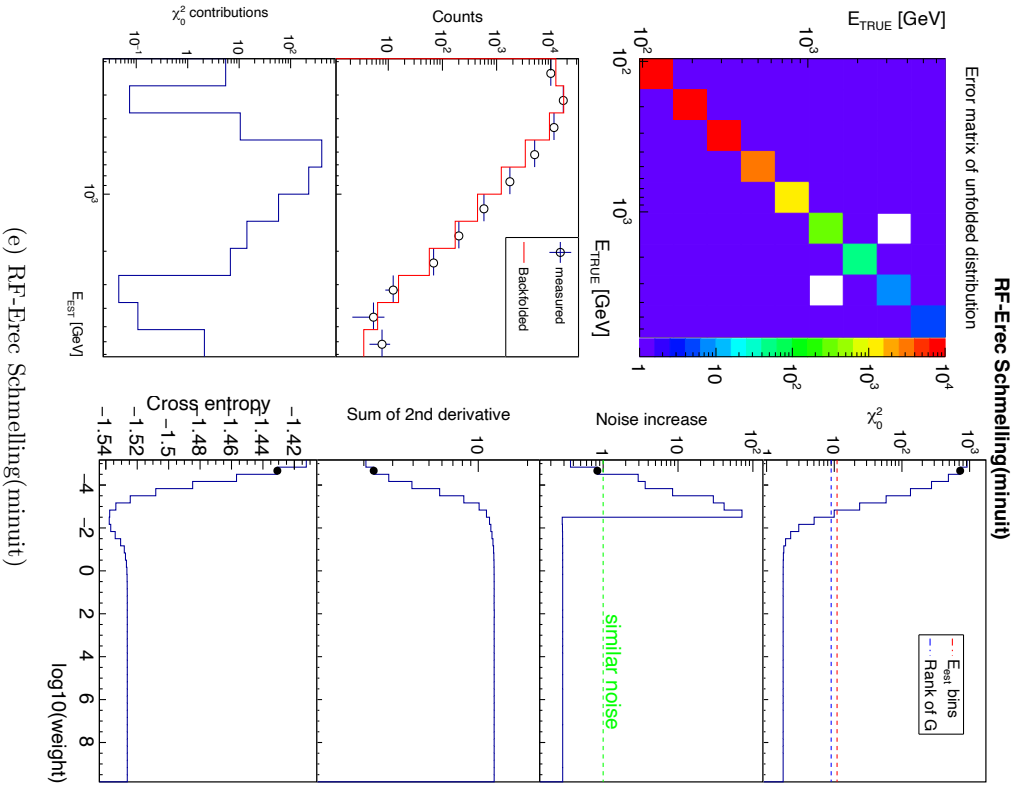
(a) RF-Erec Tikhonov

(b) LUTs-Erec Tikhonov

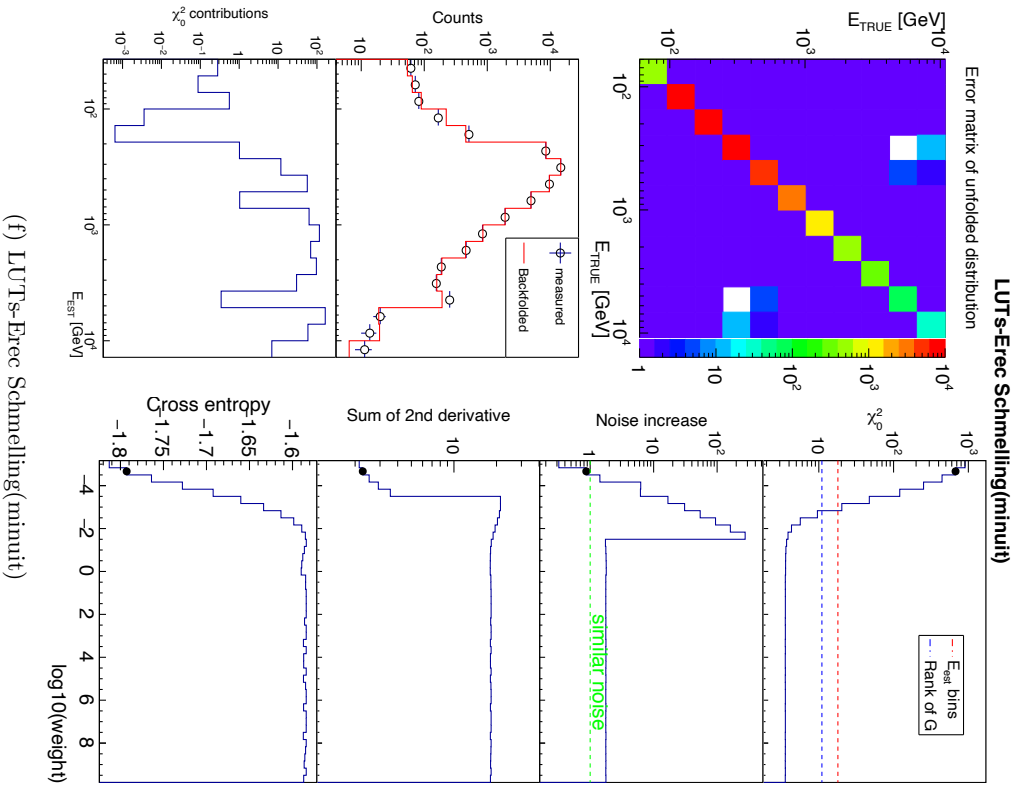


(c) RF-Erec Schmelling

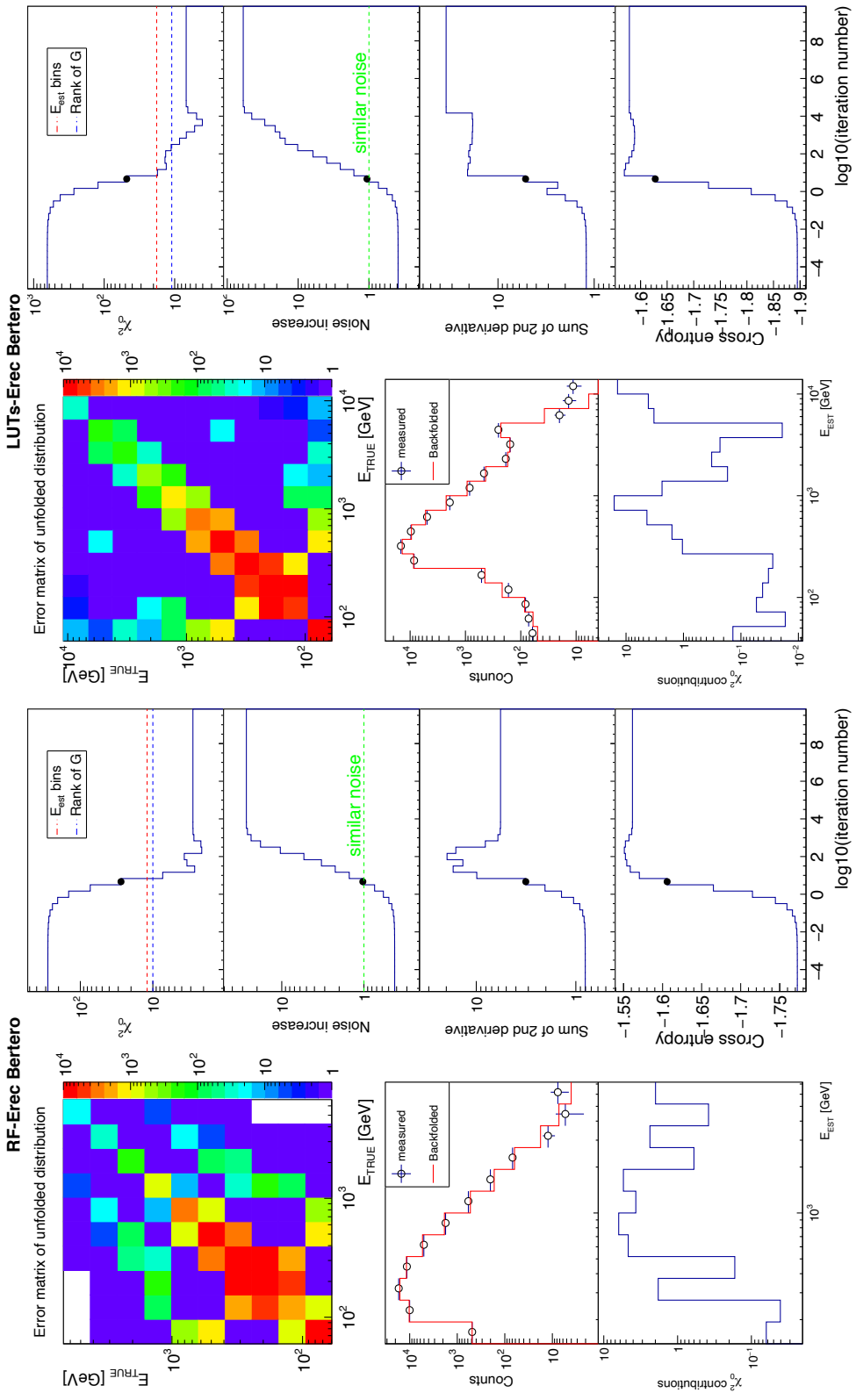
(d) LUTs-Erec Schmelling



(e) RF-Erec Schmelling(minuit)

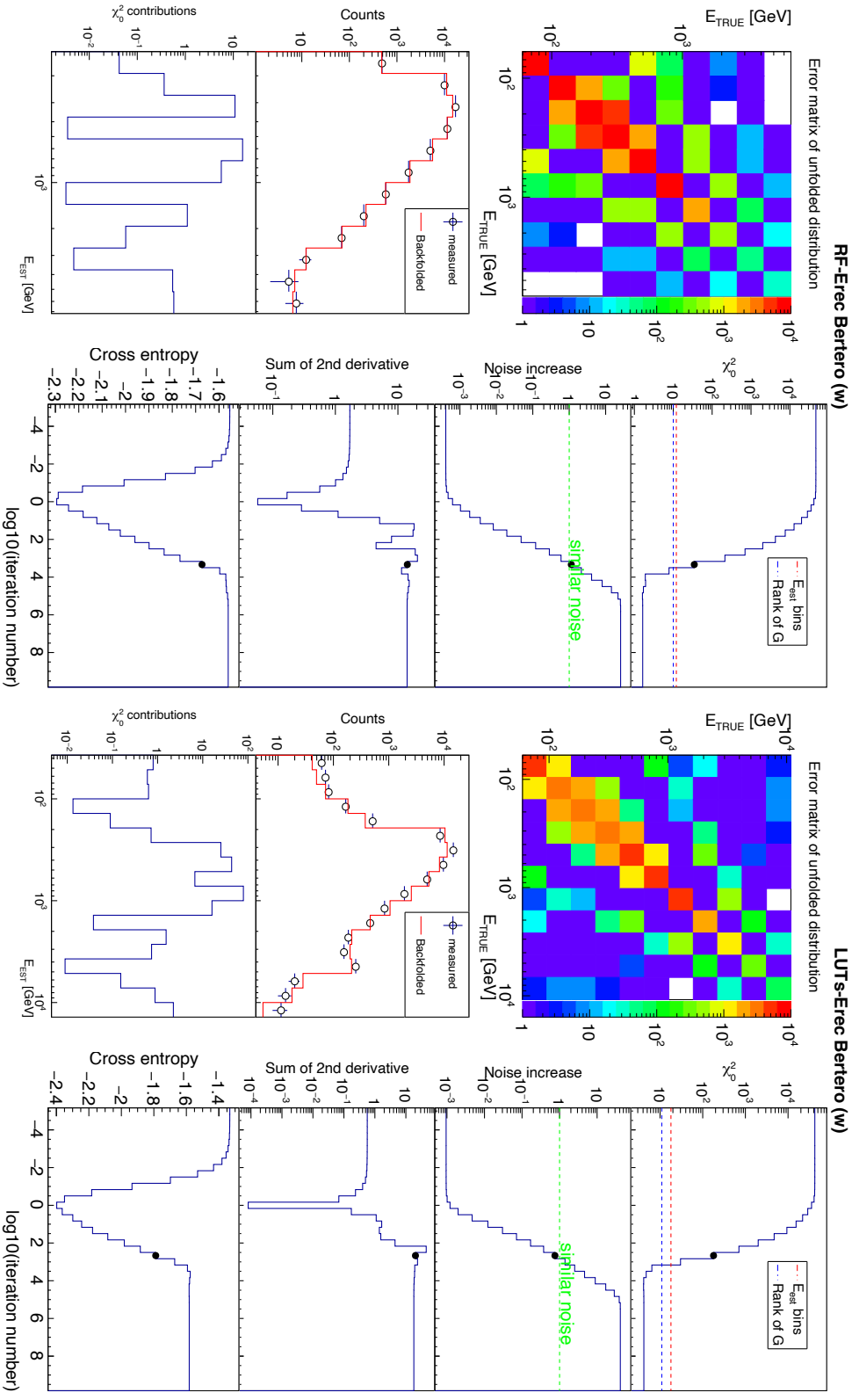


(f) LUTs-Erec Schmelling(minuit)



(g) RF-Erec Bertero

(h) LUTs-Erec Bertero



(i) RF-Erec Bertero(w)

(j) LUTs-Erec Bertero(w)

Figure H.1: Control plots for the unfoldings for the simulated observation data

H.4 Control plots in the analysis of GRB190114C data

To obtain the result of unfolding, shown in figure 5.6, the estimated energy range and true energy range are selected as shown in figure 5.3. Figure H.2 is the control plots for the results, in which the best regularization strength is determined by the point of comparable noise level.

Tikhonov

In both RF-Erec and LUTs-Erec, the solution by the criterion of keeping the noise level resulted in χ_0^2 of above 30, while the number of estimated energy bins used for migration was less than 10. Therefore the χ_0^2 are too high and I adopted the criterion of the best regularization strength to be at the point with χ_0^2 closest to the number of bins for the migration. It is natural that the noise level increases when the regularization strength with smaller χ_0^2 is selected. And the unfolding with RF-Erec had reasonable increase of noise level to about 10. In LUT, the tendency looks similar but the noise level develops in peculiar way. In the change of the regularization strength, there appears suddenly the low noise solution. It is too low and further investigation is needed what is actually happening.

Comparing the unfolded spectra among different criteria for the selection of best regularization strength, RF-Erec resulted in the spectra consistent with each other. On the other hand, LUT-Erec showed less consistent spectra, especially between the criterion of selecting comparable noise level and the others.

When χ_0^2 are compared between RF-Erec and LUTs-Erec in the same criterion of choosing the best regularization strength, RF-Erec always gives better value, although the difference is just around 1.

Schmelling

Similarly to the case for the simulated observation data, the unfolding with LUTs-Erec for the GRB data has the problem in convergence in Schmelling regularization. The Gauss-Newton iteration fails to converge in the weights above 100. The range of weight facing the failure of convergence is much wider than for the simulated observation data.

Apart from this, the control plots show normal behaviors, where χ_0^2 increases as the regularization strength lowers (the weight ω decreases), while the noise increases. It is difficult in balancing χ_0^2 and noise level. In RF-Erec, the best balance between chi_0^2 and the noise level was obtained with the criterion of searching for comparable noise level. With the noise level about 1, χ_0^2 stays around 11. In LUTs-Erec, the solution by the same criterion gives χ_0^2 about 27, which is too high. Therefore the criterion adopted is to search for χ_0^2 closest to the number of bins for the migration. In this case, the noise level becomes about 7.

Comparing the unfolded spectra among different criteria for the selection of best regularization strength, RF-Erec resulted in the spectra consistent with each other. On the other hand, the spectra with LUT-Erec showed inconsistent spectra, especially between the criterion of selecting comparable noise level (The

spectrum extends until 700 GeV) and the others (The spectra extend until above 2 TeV).

When χ_0^2 are compared between RF-Erec and LUTs-Erec in the same criterion of choosing the best regularization strength, RF-Erec always gives significantly better value. In case of Schmelling method, there is the effective number of degrees of freedom available, and RF-Erec always gives larger number than LUTs-Erec. These indicate the unfolded spectra with RF-Erec is more consistent with the measured distribution than with LUTs-Erec.

When χ_0^2 and noise level for the best unfolded spectra are compared between RF-Erec and LUTs-Erec, RF-Erec is better because the noise level is much lower while χ_0^2 is moderately worse.

Schmelling(m)

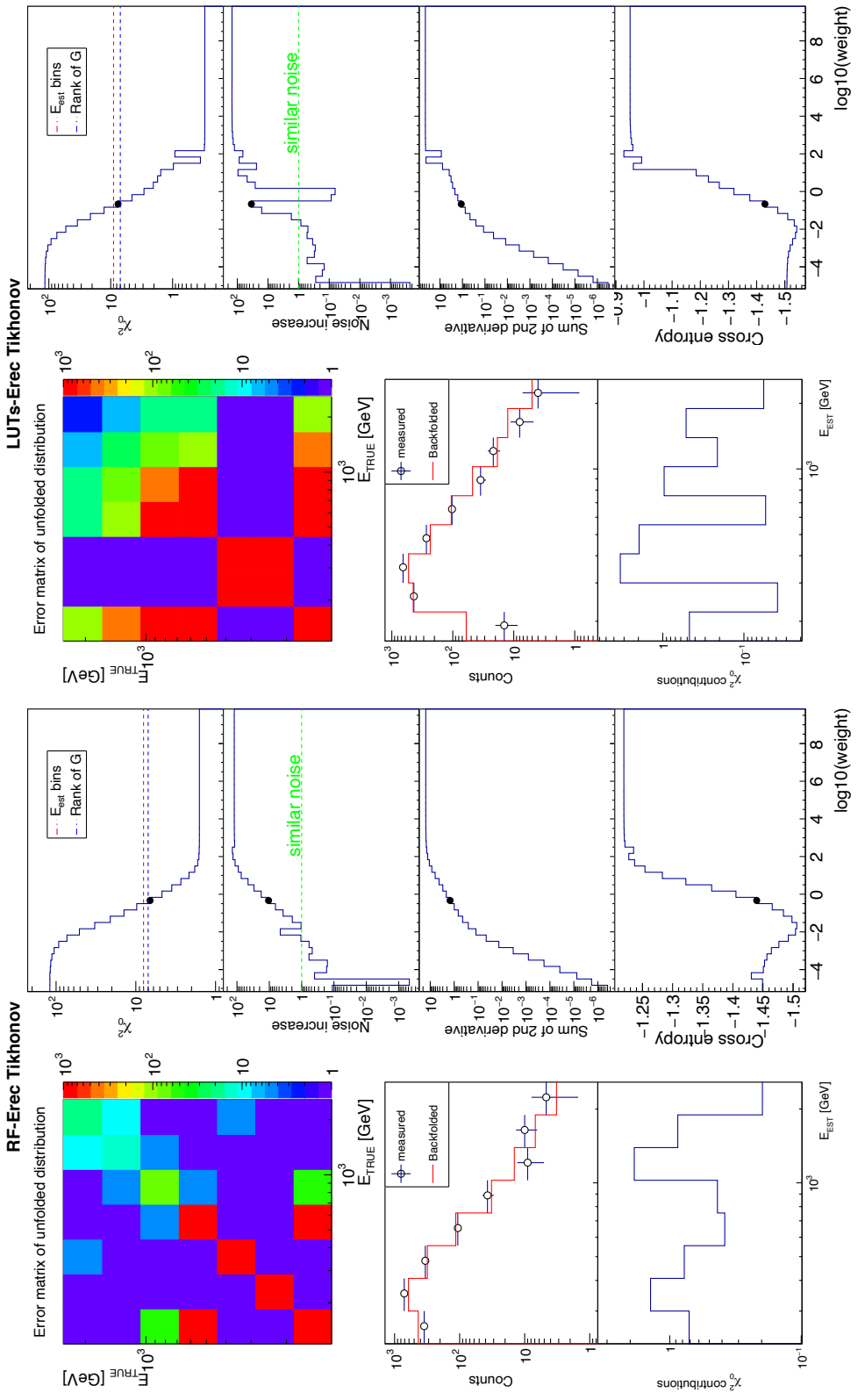
In both RF-Erec and LUTs-Erec, the solution by the criterion of keeping the noise level resulted in high χ_0^2 (about 20 and about 30 respectively), while the number of estimated energy bins used for migration was less than 10. Therefore I adopted the criterion of the best regularization strength to be at the point with χ_0^2 closest to the number of bins for the migration. In the case, the χ_0^2 become 6.6 and 9.3 respectively. Therefore, when χ_0^2 are compared between RF-Erec and LUTs-Erec in the same criterion of choosing the best regularization strength, RF-Erec gives significantly better value in those two criteria. In the criterion being at the point with χ_0^2 closest to the rank of G, χ_0^2 of RF-Erec is slightly worse than LUTs-Erec (6.6 and 5.4 respectively).

In both RF-Erec and LUTs-Erec, the unfolded spectra are consistent with each other among different criteria for the selection of best regularization strength.

Bertero and Bertero(W)

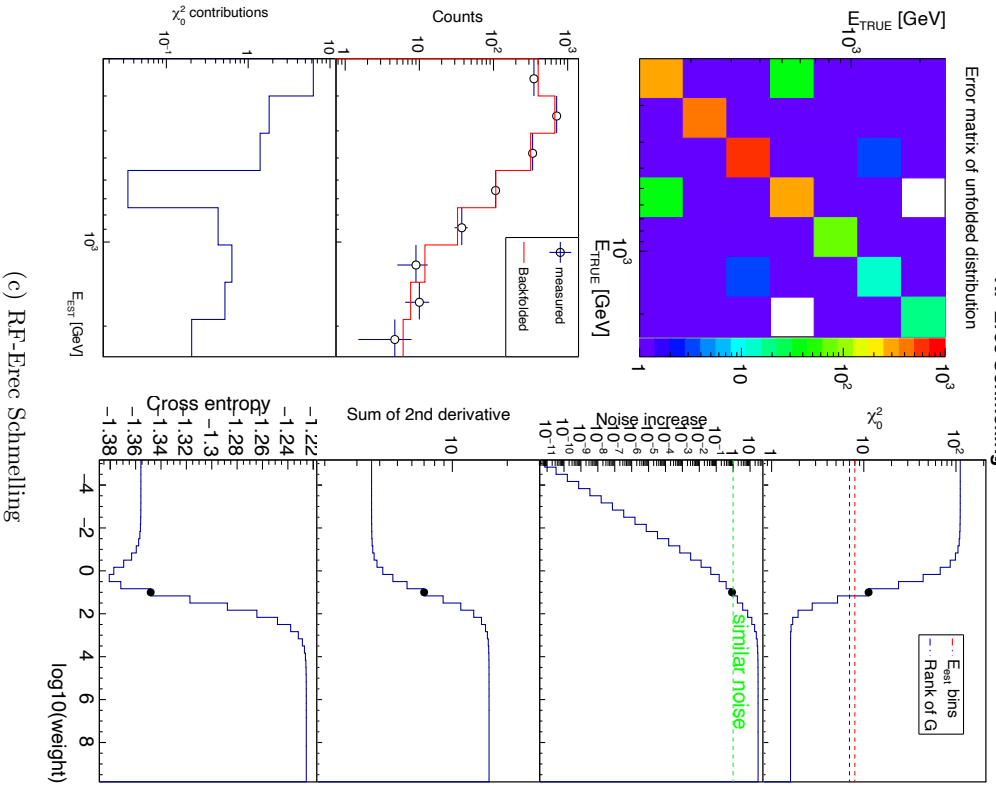
In both RF-Erec and LUTs-Erec, the solutions by the different criteria of choosing regularization strength are close to each other. Namely, the χ_0^2 and noise level are close to or the same as each other. χ_0^2 is several and at most less than 10, and noise level is about 1. As the representative unfolded spectrum I choose the one with the criterion of selecting comparable noise level.

Since the selected strengths of regularization are almost the same, the unfolded spectra are close to each other among different criteria for the selection of best regularization strength. However, in the unfolding with LUTs-Erec, the unfolded spectra by BerteroW contain a gap between the highest energy point above 1TeV and the lower energy side below $\sim$ 700 GeV.

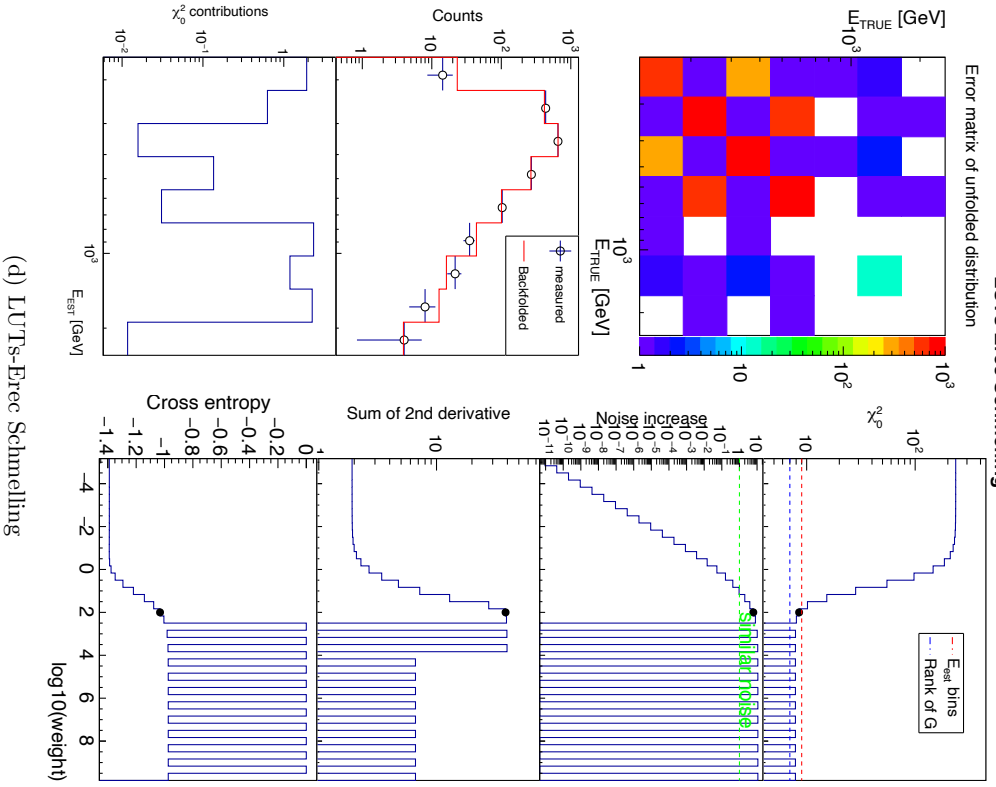


(a) RF-Erec Tikhonov

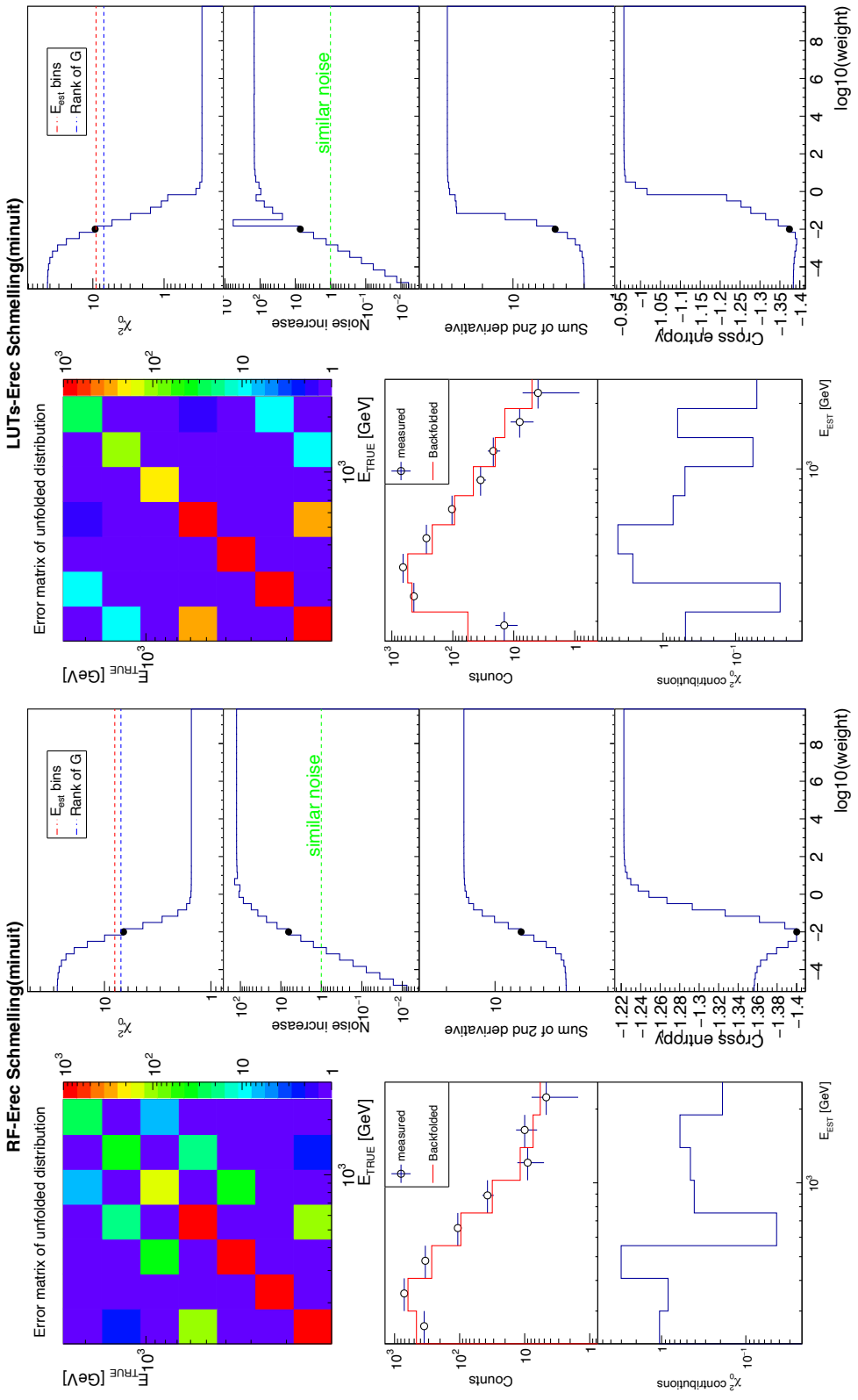
(b) LUTs-Erec Tikhonov



(c) RF-Erec Schmelling

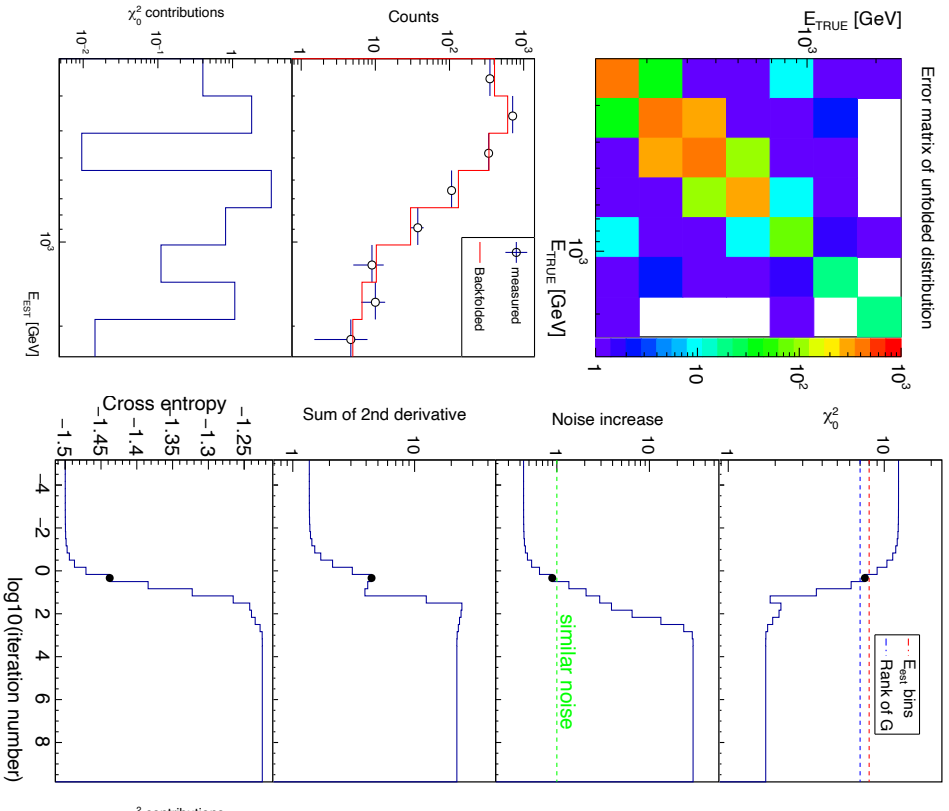


(d) LUTs-Erec Schmelling

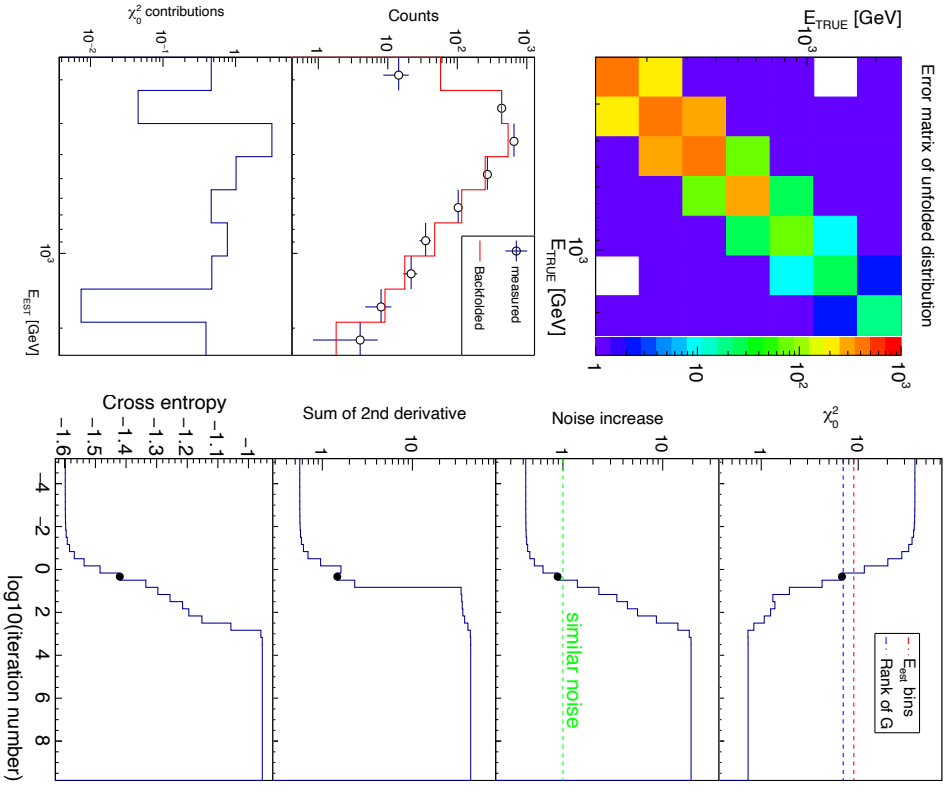


(e) RF-Erec Schmelling(minuit)

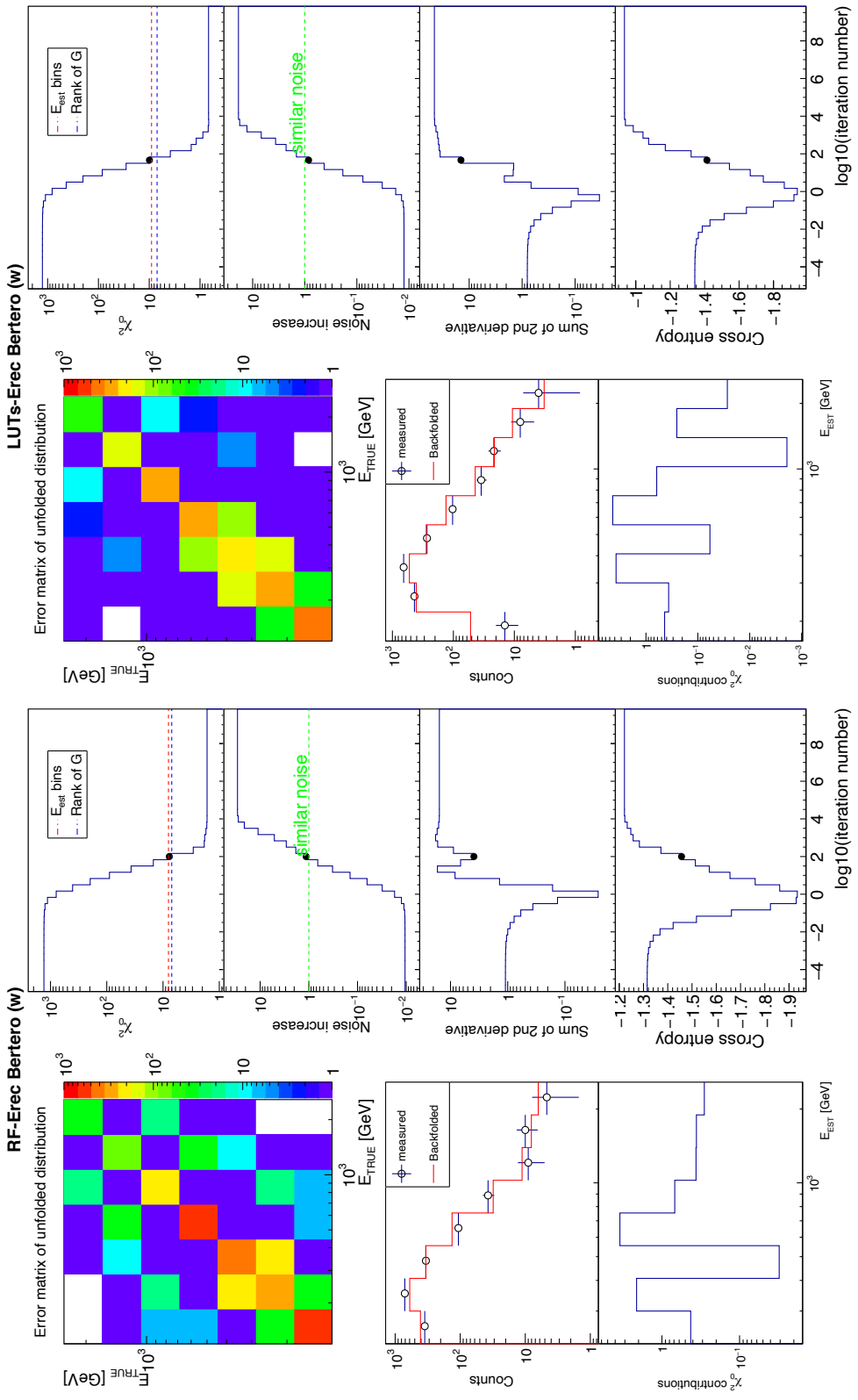
(f) LUTs-Erec Schmelling(minuit)



(g) RF-Erec Bertero



(h) LUTs-Erec Bertero



(i) LUTs-Erec Bertero(w)

(j) RF-Erec Bertero(w)

Figure H.2: Control plots for the unfoldings for the GRB 190114C data

Appendix I

The alternative analysis and the GRB team analysis

In this chapter, I compare my analysis done with RF-Erec-v5 (Alternative analysis), with that by the GRB analysis team done with RF-Erec-v3 (GRB team analysis). The difference between these two analyses is summarized in section 5.3.2.

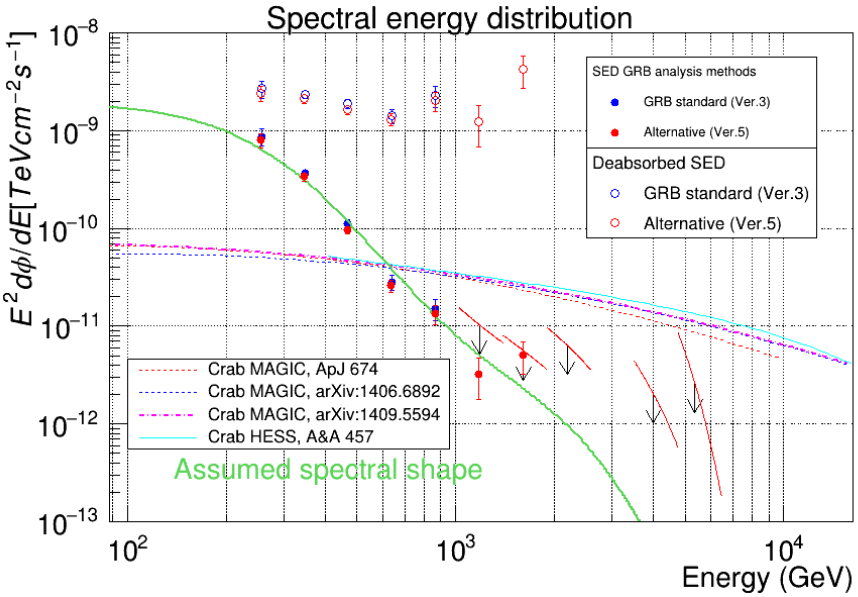


Figure I.1: The spectra of GRB 190114C by the GRB team analysis and the alternative analysis

The GRB spectra are drawn by GRB team analysis (blue) and the alternative analysis (red). he solid line is the assumed spectral shape as observed spectrum, namely the shape after the EBL absorption is applied to the power-law intrinsic spectrum. The filled circles are the observed spectral data points and the open circles are the intrinsic spectral points.

The SEDs by the different two analyses are shown in figure I.1. Both of the spectral data points are based on the same spectral shape as assumed intrinsic spectrum, which is set to be a power law with index -2.67 at the red shift of z =0.4245. The assumed observed spectrum is computed from the intrinsic spectrum applying the EBL absorption by Domínguez et. al, and it is convoluted with the collection area and the correspondent migration matrix, which are obtained using MC test data, to obtain the collection area directly for the excess events distribution versus estimated energy. Since the assumed spectrum is set to be the same, the spectral data points can be evaluated whether they are consistent with each other or not.

The overall shape of the SED are consistent with each other, therefore the scientific discussion on the source in section 5.5 is valid for both results. However the difference can still be seen above 1 TeV, where only the alternative analysis show two additional data points. It is possible to claim that the spectrum is extended until about 2 TeV.

The difference of the results is consistent with the difference in the strategy of the analysis. The source is so bright that it allows us to relax the cut because of higher signal to noise ratio. It is beneficial especially in the highest energy which always suffer from poor statistics. Therefore in the alternative analysis, the cut condition was relaxed to enlarge the collection area. The cut condition which is efficient for gaining statistics in the high energy range is θ^2 [1]. In addition, the RF-Erec-v5 can also contribute to gaining statistics, because of better direction reconstruction especially in high energy range.

However, the cut condition in the alternative analysis should be considered carefully not to introduce any systematic effect. In the lowest energy, it contains higher risk of systematic effect due to the very steep spectrum, where the changes of flux and collection area are both dramatic as a function of energy. Therefore the change of collection area as a function of energy is mitigated by relaxing Size. In the lowest energy range θ^2 cut also needs to be watched carefully, because the cut is efficiency-base and the lower energy has the larger cut value for relaxing θ^2 cut and Size cut. The largest allowed value is 0.08 deg^2, not to overlap between the on region and off regions. In the end the spectral data points of the alternative analysis become reliable especially at both lowest and highest energies.

As shown in figure I.2, the collection area in alternative is larger due to the relaxed cut condition and the improved Disp estimation by RF-Erec-v5. The effects of individual cut conditions are energy dependent, however the overall effect is more or less constantly around several tens of %. When the number of events in each energy bin is investigated, the number of excess events are increased at more or less constant ratio. At highest energy of the spectrum, the increase is dramatic. However, the increase in the background region is not peculiar, and still the number of events in the bins above 1 TeV is of the order of 1. Therefore the additional events at above 1 TeV in the alternative analysis would be genuine.

[1] The cut condition which gives a major difference of collection area between the GRB team analysis and the alternative analysis is EnergyRMS. However it is not usual to use EnergyRMS for a cut condition, therefore it is not mentioned here.

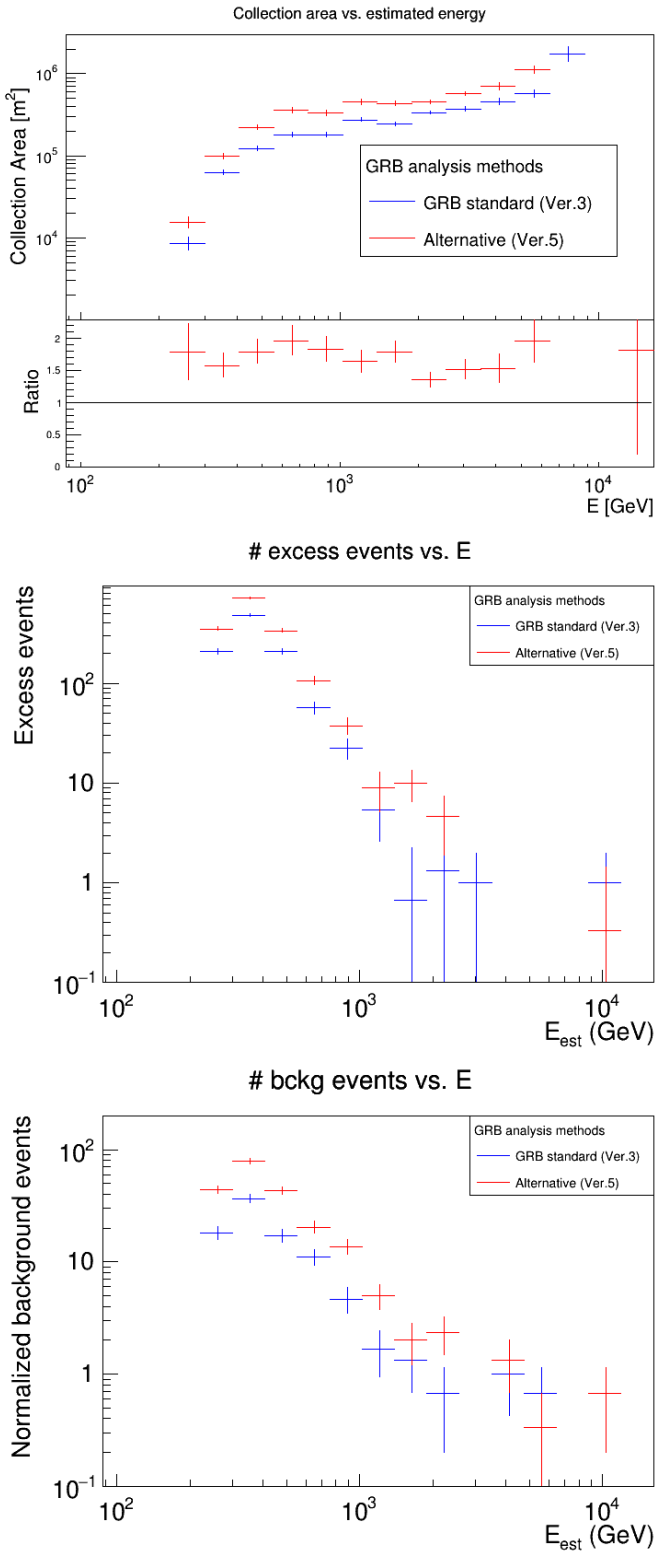


Figure I.2: The intermediate plots for the spectrum of the GRB 190114C , by the GRB team analysis and the alternative analysis

From top to bottom, the collection area, the number of number of excess events, and the number of background events, as a function of energy.

List of Figures

List of Tables

List of acronyms

CGRO	Compton Gamma Ray Observatory.
AAS	Automatic Alert System.
AGN	Active Galactic Nuclei.
ANN	Artificial Neural Networks.
Az	Azimuth angle.
BAGGing	Bootstrap AGGregatING.
Bagging	Bootstrap Aggregating.
BAT	Burst Alert Telescope.
BATSE	Burst And Transient Source Experiment.
BDT	Boosted Decision Tree.
BH	Black Hole.
CMB	Cosmic Microwave Background.
CORSIKA	COsmic Ray SImulation for KASCADE.
DC	Davies-Cotton.
DM	Dark Matter.
DT	Discriminator Threshold (in the context of the observation setting).
DT	Decision Tree.
EAS	Extensive Air Shower.
EBL	Extragalactic Background Light.
FADC	Fast Analog to Digital Converter.
FoV	Field of View.
GBM	Gamma-ray Burst Monitor.
GRB	Gamma-Ray Burst.
IACT	Imaging Atmospheric Cherenkov Telescope.
IC	Inverse Compton process.
LAT	Large-Area Telescope.
LIV	Lorentz Invariance Violation.

LUT	Look-Up-Table.
LUTs-Erec	Look-Up Tables based Energy reconstruction.
MAGIC	Major Atmospheric Gamma Imaging Cherenkov telescope system.
MARS	MAGIC Analysis and Reconstruction Software.
MC	Monte Carlo simulation.
MIP	A minimum ionizing particle.
ML	Machine Learning.
MMCS	MAGIC Monte Carlo Software.
MSE	Mean Squared Error.
NDF	Number of Degree of Freedom.
NERMS	the EnergyRMS normalised with the estimated energy.
NIR	Near-InfraRed.
NKG	Nishimura-Kamata-Greisen.
NN	Next-Neighbor pixels.
NS	Neutron Star.
NSB	Night Sky Background.
OOB	Out-Of-Bag.
ORM	the Roque de los Muchachos Observatory.
phe	photoelectron.
PMT	Photo Multiplier Tube.
PSF	Point Spread Function.
PWN	Pulsar Wind Nebula.
RF	Random Forest.
RF-Erec	Random Forest based Energy reconstruction.
RSS	Residual Sum of Squares.
SD	Standard Deviation.
SED	Spectral Energy Distribution.
SN	Super Nova.
SNR	Super Nova Remnant.
SSC	Synchrotron-Self Compton process.
UV	Ultra Violet.
VHE	Very High Energy γ-rays ($E \gtrsim 50$ GeV).
WC	Winston Cone.
Zd	the angular distance from the source to the zenith (Zenith Angle,ZA).

List of symbols

GRB physics (section 5.5 and G.2)

Symbol	Description
N_e	the number of electrons
ϵ_e and ϵ_B	the fraction of the shock internal energy that is partitioned to electrons, and the one to magnetic fields
γ_m and γ_M	the minimum and maximum Lorentz factors of the electron energy distribution, respectively
$\hbar$	$= h/2\pi$, reduced Planck constant
e_1 and e_2	the internal energy in upstream and downstream
h	Planck constant, $6.626 \times 10^{-27}\,\mathrm{erg\,s}$
n_1 and n_2	the number density in upstream and downstream
t_L	Larmor time
A_*	a parameter characterizing the normalization of the density
B	magnetic field
P	radiated power
R	the distance from a GRB central engine
Γ	Bulk Lorentz factor (the bulk motion of the shock wave)
α	pitch angle (the angle between a particle and the magnetic field in the shock wave)
β	relative speed to the light c
γ_c	Lorentz factor correspondent to ν_c, called cooling energy or cooling Lorentz factor
γ_e	Electron Lorentz factor
γ	Lorentz factor
ν_M	The cutoff in synchrotron spectrum correspondent to the maximum Lorentz factor of the electron injection γ_M
ν_c	The spectral break in the synchrotron spectrum between ν_m and ν_M due to cooling
ν_m	The frequency in synchrotron spectrum correspondent to the minimum Lorentz factor of the electron injection γ_m
ν_{ch}	critical frequency of synchrotron emission
$\nu_{\mathrm{peak}}^{\mathrm{SSC}}$ and $\varepsilon_{\mathrm{peak}}^{\mathrm{SSC}}$	peak of SSC emission, at the comoving frame of the shock wave (in the unit of frequency) and the frame of an observer (in the unit of energy), respectively

Symbol	Description
$\nu_{\mathrm{peak}}^{\mathrm{Syn}}$ and $\varepsilon_{\mathrm{peak}}^{\mathrm{Syn}}$	peak of synchrotron emission, at the comoving frame of the shock wave (in the unit of frequency) and the frame of an observer (in the unit of energy), respectively
ν	frequency of photon
σ_T	Thomson cross section, $6.652 \times 10^{-25}\,\mathrm{cm}^2$
c	speed of light, 2992458 m/s
e	electric charge of an election, $1.602 \times 10^{-19}\,\mathrm{C}$
m_e	electron mass, $0.511\ \mathrm{MeV}/c^2$
m_p	proton mass, $938\ \mathrm{MeV}/c^2$
n	the number density of circumburst medium
p	Spectral index of the energy distribution of electrons injected to the shock wave
s	Power-law index for the radial profile of circumburst medium
t_{cool}	synchrotron cooling time

Unfolding (section 3.11.1, 4.6.4, 5.4.2 and appendix H)

Symbol	Description
G	Gram's matrix
K	covariance matrix for a measured distribution
M	migration matrix
T	covariance matrix for an unfolded distribution
χ^2	chi-square of the unfolding process (regularization term is added to χ_0^2)
χ_0^2	chi-square of the unfolding process (between the measured distribution and the backfolded distribution)
ω	a weight to steer the regularization strength
σ_y	error of the measured distribution
$\mathbf{y}$ and $\mathbf{s}$	the signal distribution over energy, before and after the unfolding respectively
i	iteration number for Bertero's method, to steer the regularization strength
n_r	the rank of G (the number of eigenvectors)
n_y and n_s	the number of energy bins, before and after the unfolding respectively

Bibliography

[1] B. P. Abbott et al. "Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A". In: *Astrophys. J.* 848.2 (Oct. 2017), p. L13. ISSN: 2041-8213. DOI: 10.3847/2041-8213/aa920c. arXiv: 1710.05834.

[2] B. P. Abbott et al. "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". In: *Phys. Rev. Lett.* 119.16 (Oct. 2017), p. 161101. ISSN: 10797114. DOI: 10.1103/PhysRevLett.119.161101.

[3] A. A. Abdo et al. "FERMI OBSERVATIONS OF GRB 090902B: A DISTINCT SPECTRAL COMPONENT IN THE PROMPT AND DELAYED EMISSION". In: *Astrophys. J. Lett.* 706.1 (2009), p. L138. ISSN: 1538-4357. DOI: 10.1088/0004-637X/706/1/L138.

[4] A. A. Abdo et al. "Gamma-ray flares from the Crab Nebula". In: *Science (80-.).* 331.6018 (Feb. 2011), pp. 739–742. ISSN: 10959203. DOI: 10.1126/science.1199705. arXiv: 1011.3855.

[5] A. A. Abdo et al. "Insights into the high-energy γ-ray emission of Markarian 501 from extensive multifrequency observations in the Fermi era". In: *Astrophys. J.* 727.2 (Feb. 2011), p. 129. ISSN: 15384357. DOI: 10.1088/0004-637X/727/2/129.

[6] J Abraham et al. *A Study of the Effect of Molecular and Aerosol Conditions in the Atmosphere on Air Fluorescence Measurements at the Pierre Auger Observatory*. Tech. rep. 2010. arXiv: 1002.0366v1.

[7] V. A. Acciari et al. "Observation of inverse Compton emission from a long γ-ray burst". In: *Nature* 575.7783 (Nov. 2019), pp. 459–463. ISSN: 14764687. DOI: 10.1038/s41586-019-1754-6.

[8] V. A. Acciari et al. "Teraelectronvolt emission from the γ-ray burst GRB 190114C". In: *Nature* 575.7783 (Nov. 2019), pp. 455–458. ISSN: 14764687. DOI: 10.1038/s41586-019-1750-x.

[9] M. Ackermann et al. "Fermi-LAT observations of the gamma-ray burst GRB 130427A". In: *Science (80-.).* 343.6166 (Jan. 2014), pp. 42–47. ISSN: 10959203. DOI: 10.1126/science.1242353.

[10] F. Aharonian et al. "High energy astrophysics with ground-based gamma ray detectors". In: *Reports Prog. Phys.* (2008). ISSN: 00344885. DOI: 10.1088/0034-4885/71/9/096901.

[11] M. L. Ahnen et al. “Performance of the MAGIC telescopes under moonlight”. In: *Astropart. Phys.* 94 (Sept. 2017), pp. 29–41. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2017.08.001. arXiv: 1709.00283.

[12] M.L. L. Ahnen et al. “A cut-off in the TeV gamma-ray spectrum of the SNR Cassiopeia A”. In: *Mon. Not. R. Astron. Soc.* 472.3 (Dec. 2017), pp. 2956–2962. DOI: 10.1093/mnras/stx2079. URL: https://academic.oup.com/mnras/article/472/3/2956/4082233.

[13] J. Albert et al. “Implementation of the Random Forest method for the Imaging Atmospheric Cherenkov Telescope MAGIC”. In: *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* 588.3 (2008), pp. 424–432. ISSN: 01689002. DOI: 10.1016/j.nima.2007.11.068.

[14] J. Albert et al. “Probing quantum gravity using photons from a flare of the active galactic nucleus Markarian 501 observed by the MAGIC telescope”. In: *Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys.* 668.4 (Oct. 2008), pp. 253–257. ISSN: 03702693. DOI: 10.1016/j.physletb.2008.08.053. arXiv: 0708.2889. URL: https://ui.adsabs.harvard.edu/abs/2008PhLB..668..253M/abstract.

[15] J. Albert et al. “Unfolding of differential energy spectra in the MAGIC experiment”. In: *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* 583.2-3 (Dec. 2007), pp. 494–506. ISSN: 01689002. DOI: 10.1016/j.nima.2007.09.048. arXiv: 0707.2453.

[16] J. Albert et al. “VHE γ-Ray Observation of the Crab Nebula and its Pulsar with the MAGIC Telescope”. In: *Astrophys. J.* 674.2 (Feb. 2008), pp. 1037–1055. ISSN: 0004-637X. DOI: 10.1086/525270.

[17] J. Aleksić et al. “Performance of the MAGIC stereo system obtained with Crab Nebula data”. In: *Astropart. Phys.* 35.7 (Feb. 2012), pp. 435–448. ISSN: 0927-6505. DOI: 10.1016/J.ASTROPARTPHYS.2011.11.007. URL: https://www.sciencedirect.com/science/article/pii/S0927650511002064?via%7B%5C%%7D3Dihub.

[18] J. Aleksić et al. “The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system”. In: *Astropart. Phys.* 72 (Jan. 2016), pp. 61–75. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2015.04.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0927650515000663.

[19] J. Aleksić et al. “The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula”. In: *Astropart. Phys.* 72 (Jan. 2016), pp. 76–94. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2015.02.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0927650515000316.

[20] E. Aliu et al. “Improving the performance of the single-dish Cherenkov telescope MAGIC through the use of signal timing”. In: *Astropart. Phys.* 30.6 (2009), pp. 293–305. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2008.10.003.

[21] E. Aliu et al. "Observation of pulsed γ-rays above 25 GeV from the crab pulsar with MAGIC". In: *Science (80-.).* 322.5905 (Nov. 2008), pp. 1221–1224. ISSN: 10959203. DOI: 10.1126/science.1164718.

[22] F. Arqueros, F. Blanco, and J. Rosado. "Analysis of the fluorescence emission from atmospheric nitrogen by electron excitation, and its application to fluorescence telescopes". In: *New J. Phys.* 11.6 (June 2009), p. 065011. ISSN: 13672630. DOI: 10.1088/1367-2630/11/6/065011. URL: https://iopscience.iop.org/article/10.1088/1367-2630/11/6/065011.

[23] E. Berger. "A short gamma-ray burst "no-host" problem? Investigating large progenitor offsets for short GRBs with optical afterglows". In: *Astrophys. J.* 722.2 (Oct. 2010), pp. 1946–1961. ISSN: 15384357. DOI: 10.1088/0004-637X/722/2/1946.

[24] E. Berger, W. Fong, and R. Chornock. "An r-process kilonova associated with the short-hard GRB 130603B". In: *Astrophys. J. Lett.* 774.2 (Sept. 2013), p. L23. ISSN: 20418205. DOI: 10.1088/2041-8205/774/2/L23. arXiv: 1306.3960.

[25] Edo Berger. "Short duration gamma ray bursts". In: *Annu. Rev. Astron. Astrophys.* 52.Aug (2014), pp. 43–105. ISSN: 03044289. DOI: 10.1146/annurev-astro-081913-035926. arXiv: 1311.2603.

[26] M G Bernardini et al. "The X-ray light curve of gamma-ray bursts: clues to the central engine". In: *A&A* 539 (2012). DOI: 10.1051/0004-6361/201117895.

[27] Konrad Bernlöhr. "Impact of atmospheric parameters on the atmospheric Cherenkov technique". In: *Astropart. Phys.* 12.4 (2000), pp. 255–268. ISSN: 09276505. DOI: 10.1016/S0927-6505(99)00093-6.

[28] M. Bertero. "Linear Inverse and III-Posed Problems". In: *Adv. Electron. Electron Phys.* 75.C (Jan. 1989), pp. 1–120. ISSN: 00652539. DOI: 10.1016/S0065-2539(08)60946-4.

[29] Christopher M. Bishop. *Pattern Recogniton and Machine Learning*. Springer-Verlag New York, 2006, p. 738. ISBN: 978-0-387-31073-2. URL: https://www.springer.com/gp/book/9780387310732%20http://users.isr.ist.utl.pt/%7B~%7Dwurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%202006.pdf.

[30] R. D. Blandford and C. F. McKee. "Fluid dynamics of relativistic blast waves". In: *Phys. Fluids* 19.8 (Aug. 1976), pp. 1130–1138. ISSN: 10706631. DOI: 10.1063/1.861619. URL: https://aip.scitation.org/doi/10.1063/1.861619.

[31] J. S. Bloom et al. "The unusual afterglow of the γ-ray burst of 26 March 1998 as evidence for a supernova connection". In: *Nature* 401.6752 (Sept. 1999), pp. 453–456. ISSN: 00280836. DOI: 10.1038/46744.

[32] Gerhard Bohm and Günter Zech. *Introduction to Statistics and Data Analysis for Physicists*. 2010. ISBN: 9783935702416.

[33] Leo Breiman. "Bagging Predictors". In: *Mach. Learn.* 24.2 (1996), pp. 123–140. ISSN: 08856125. DOI: 10.1023/A:1018054314350. URL: http://link.springer.com/10.1023/A:1018054314350.

[34] Leo Breiman. *Random Forests*. Tech. rep. 2001, pp. 5–32. URL: https://link.springer.com/content/pdf/10.1023%7B%5C%%7D2FA%7B%5C%%7D3A1010933404324.pdf.

[35] Leo Breiman and Adele Cutler. *Random Forests*. URL: https://www.stat.berkeley.edu/%7B~%7Dbreiman/RandomForests/.

[36] Thomas Bretz and Daniela Dorner. "MARS-CheObs goes Monte Carlo". In: *Proc 31st ICRC, Lodz, Poland, Id. 1259, 2009* (2009). URL: https://srl.utu.fi/AuxDOC/kocharov/ICRC2009/pdf/icrc1259.pdf.

[37] Torsten Bringmann, Lars Bergström, and Joakim Edsjö. "New gamma-ray contributions to supersymmetric dark matter annihilation". In: *J. High Energy Phys.* 2008.1 (Jan. 2008), p. 049. ISSN: 11266708. DOI: 10.1088/1126-6708/2008/01/049. arXiv: 0710.3169.

[38] Rene Brun and Fons Rademakers. "ROOT - An object oriented data analysis framework". In: *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* 389.1-2 (Apr. 1997), pp. 81–86. ISSN: 01689002. DOI: 10.1016/S0168-9002(97)00048-X.

[39] A Castro-Tirado. "GRB 190114C: refined redshift by the 10.4m GTC". In: *GCN Circ.* gcn3 (2019.23708 (2019).

[40] H. E. S. S. Collaboration et al. "First Ground-based Measurement of Sub-20 GeV to 100 GeV gamma-rays from the Vela Pulsar with H.E.S.S. II". In: (July 2018). DOI: 10.1051/0004-6361/201732153. arXiv: 1807.01302. URL: http://arxiv.org/abs/1807.01302%20http://dx.doi.org/10.1051/0004-6361/201732153.

[41] S.C. Commichau et al. "Monte Carlo studies of geomagnetic field effects on the imaging air Cherenkov technique for the MAGIC telescope site". In: *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* 595.3 (Oct. 2008), pp. 572–586. ISSN: 01689002. DOI: 10.1016/j.nima.2008.07.144. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168900208010978.

[42] Charles D. Dermer, James Chiang, and Kurt E. Mitman. "Beaming, Baryon Loading, and the Synchrotron Self-Compton Component in Gamma-Ray Bursts". In: *Astrophys. J.* 537.2 (July 2000), pp. 785–795. ISSN: 0004-637X. DOI: 10.1086/309061.

[43] Thomas G. Dietterich. "Experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization". In: *Mach. Learn.* 40.2 (Aug. 2000), pp. 139–157. ISSN: 08856125. DOI: 10.1023/A:1007607513941.

[44] A. Dominguez et al. "Extragalactic Background Light Inferred from AEGIS Galaxy SED-type Fractions". In: *Mon. Not. R. Astron. Soc.* 410.4 (July 2011), pp. 2556–2578. DOI: 10.1111/j.1365-2966.2010.17631.x. arXiv: 1007.1459. URL: http://arxiv.org/abs/1007.1459%20http://dx.doi.org/10.1111/j.1365-2966.2010.17631.x.

[45] Michele Doro. "The Commissioning and Characterization of the Calibration System of the MAGIC Telescope Candidato : Michele Doro". PhD thesis. 2004.

[46] Ralph Engel, Dieter Heck, and Tanguy Pierog. "Extensive Air Showers and Hadronic Interactions at High Energy". In: *Annu. Rev. Nucl. Part. Sci.* 61.1 (Nov. 2011), pp. 467–489. ISSN: 0163-8998. DOI: 10.1146/annurev.nucl.012809.104544.

[47] Manuel Fernández-Delgado et al. *Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?* Tech. rep. 2014, pp. 3133–3181. URL: http://www.mathworks.es/products/neural-network..

[48] W. Fong et al. "Demographics of the galaxies hosting short-duration gamma-ray bursts". In: *Astrophys. J.* 769.1 (May 2013), p. 56. ISSN: 15384357. DOI: 10.1088/0004-637X/769/1/56. arXiv: 1302.3221.

[49] A. S. Fruchter et al. "Long γ-ray bursts and core-collapse supernovae have different environments". In: *Nature* 441.7092 (May 2006), pp. 463–468. ISSN: 14764687. DOI: 10.1038/nature04787. arXiv: 0603537 [astro-ph].

[50] Christian Fruck. "The Galactic Center resolved with MAGIC and a new technique for Atmospheric Calibration". PhD thesis. 2015. URL: http://mediatum.ub.tum.de/doc/1237928/1237928.pdf.

[51] Stefan Funk. "Ground-and Space-Based Gamma-Ray Astronomy". In: *Annu. Rev. Nucl. Part. Sci* 65 (2015), pp. 245–77. DOI: 10.1146/annurev-nucl-102014-022036. URL: www.annualreviews.org.

[52] Thomas K. Gaisser, Ralph Engel, and Elisa Resconi. *Cosmic rays and particle physics*. Cambridge University Press, June 2016, pp. 1–444. ISBN: 9781139192194. DOI: 10.1017/CBO9781139192194.

[53] T. J. Galama et al. "An unusual supernova in the error box of the γ-ray burst of 25 April 1998". In: *Nature* 395.6703 (Oct. 1998), pp. 670–672. ISSN: 00280836. DOI: 10.1038/27150.

[54] He Gao et al. "A complete reference of the analytical synchrotron external shock models of gamma-ray bursts". In: *New Astron. Rev.* 57.6 (Dec. 2013), pp. 141–190. ISSN: 13876473. DOI: 10.1016/j.newar.2013.10.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1387647313000171.

[55] He Gao et al. "Searching for Magnetar-powered Merger-novae from Short GRBS". In: *Astrophys. J.* 837.1 (Mar. 2017), p. 50. ISSN: 1538-4357. DOI: 10.3847/1538-4357/aa5be3. arXiv: 1608.03375.

[56] Markus Gaug, Lluís Font, and Camilla Maggio. "Site Characterization of the Northern Site of the Cherenkov Telescope Array". In: *EPJ Web Conf.* Vol. 144. EDP Sciences, May 2017. DOI: 10.1051/epjconf/201714401010.

[57] Benyamin Ghojogh and Mark Crowley. *The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial*. Tech. rep. 2019. arXiv: 1905.12787. URL: http://arxiv.org/abs/1905.12787.

[58] J. F. Graham and A. S. Fruchter. "The metal aversion of long-duration gamma-ray bursts". In: *Astrophys. J.* 774.2 (Sept. 2013), p. 119. ISSN: 15384357. DOI: 10.1088/0004-637X/774/2/119.

[59] Peter K. F. Grieder. *Exentsive Air Showers and High Energy Phenomena*. Springer Berlin Heidelberg, 2010. DOI: 10.1007/978-3-540-76941-5.

[60] J. D. Gropp. "GRB 190114C: Swift detection of a very bright burst with a bright optical counterpart". In: *GCN Circ.* 15.gcn3 (2019) (2019). URL: https://gcn.gsfc.nasa.gov/gcn3/23688.

[61] Daniel Alberto Guberman. "MAGIC observations with bright Moon and their application to measuring the VHE gamma-ray spectral cut-off of the PeVatron candidate Cassiopeia A Director : Director :" PhD thesis. 2018.

[62] R Hamburg. "GRB 190114C: Fermi GBM detection". In: *GCN Circ.* gcn3 (2019.23707 (2019).

[63] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. "The element of statistical learning". In: *Springer Ser. Stat.* (2008).

[64] D Heck and T Pierog. *Extensive Air Shower Simulation with CORSIKA : A User's Guide (Version CORSIKA 6.99x from August 8, 2011)*. 2011. URL: http://www-ik.fzk.de/corsika/physics%7B%5C_%7Ddescription/corsika%7B%5C_%7Dphys.html.

[65] Thomas Hengstebeck. "Measurement of the energy spectrum of the BL Lac object PG1553+113 with the MAGIC Telescope in 2005 and 2006". PhD thesis. 2006.

[66] A. M. Hillas. "Cerenkov light images of EAS produced by primary gamma". In: *NASA. Goddard Sp. Flight Cent. 19th Intern. Cosm. Ray Conf., Vol. 3 p 445-448 (SEE N85-34862 23-93)* 3 (1985).

[67] W. Hofmann et al. "Comparison of techniques to reconstruct VHE gamma-ray showers from multiple stereoscopic Cherenkov images". In: *Astropart. Phys.* 12.3 (1999), pp. 135–143. ISSN: 09276505. DOI: 10.1016/S0927-6505(99)00084-5.

[68] J. H. Hubbell, H. A. Gimm, and I. Overbo. "Pair, Triplet, and Total Atomic Cross Sections (and Mass Attenuation Coefficients) for 1 MeV-100 GeV Photons in Elements Z=1 to 100". In: *J. Phys. Chem. Ref. Data* 9.4 (1980), pp. 1023–1148. ISSN: 15297845. DOI: 10.1063/1.555629.

[69] Hemant Ishwaran. "The effect of splitting on random forests". In: *Mach. Learn.* (2015). ISSN: 15730565. DOI: 10.1007/s10994-014-5451-2.

[70] John David Jackson. *Classical electrodynamics*. 3rd ed. New York, {NY}: Wiley, 1999. ISBN: 9780471309321. URL: http://cdsweb.cern.ch/record/490457.

[71] Gareth James et al. *An Introduction to Statistical Learning*. Vol. 103. 2013. ISBN: 978-1-4614-7137-0. DOI: 10.1007/978-1-4614-7138-7. URL: http://link.springer.com/10.1007/978-1-4614-7138-7.

[72] J. I. Katz. "Two populations and models of gamma-ray bursts". In: *Astrophys. J.* 422 (Feb. 1994), p. 248. ISSN: 0004-637X. DOI: 10.1086/173723.

[73] Shiho Kobayashi. "Light Curves of Gamma-Ray Burst Optical Flashes". In: *Astrophys. J.* 545.2 (Dec. 2000), pp. 807–812. ISSN: 0004-637X. DOI: 10.1086/317869.

[74] Chryssa Kouveliotou et al. "Identification of two classes of gamma-ray bursts". In: *Astrophys. J.* 413 (Aug. 1993), p. L101. ISSN: 0004-637X. DOI: 10.1086/186969.

[75] Julian D.G. Krause. "Resolved gamma ray emission of the supernova remnant W51C and HESS J1857+026 obtained with the MAGIC telescopes". PhD thesis. 2012, p. 228. URL: http://inspirehep.net/record/1419574?ln=en.

[76] H. A. et al. Krimm. "GRB 190114C: Swift-BAT refined analysis." In: *GCN Circ.* gcn3 (2019.23724 (2019).

[77] Pawan Kumar and Bing Zhang. *The physics of gamma-ray bursts & relativistic jets*. 2015. DOI: 10.1016/j.physrep.2014.09.008.

[78] Antoine Letessier-Selvon and Todor Stanev. "Ultrahigh energy cosmic rays". In: *Rev. Mod. Phys.* 83.3 (Sept. 2011), pp. 907–942. ISSN: 0034-6861. DOI: 10.1103/RevModPhys.83.907. URL: https://link.aps.org/doi/10.1103/RevModPhys.83.907.

[79] T.-P. Li and Y.-Q. Ma. "Analysis methods for results in gamma-ray astronomy". In: *Astrophys. J.* 272 (Sept. 1983), p. 317. ISSN: 0004-637X. DOI: 10.1086/161295. URL: http://adsabs.harvard.edu/doi/10.1086/161295.

[80] Malcolm S. Longair and Malcolm S. "High Energy Astrophysics". In: *High Energy Astrophys. by Malcolm S. Longair, Cambridge, UK Cambridge Univ. Press. 2011* (2011).

[81] Rubén López Coto. "Very-high-energy Gamma-ray Observations of Pulsar Wind Nebulae and Cataclysmic Variable Stars with MAGIC and Development of Trigger Systems for IACTs". PhD thesis. 2017. ISBN: 978-3-319-44750-6. DOI: 10.1007/978-3-319-44751-3. URL: http://link.springer.com/10.1007/978-3-319-44751-3.

[82] Eckart Lorenz. "Status of the 17 m diameter MAGIC telescope". In: *New Astron. Rev.* 48.5-6 (Apr. 2004), pp. 339–344. ISSN: 13876473. DOI: 10.1016/j.newar.2003.12.059.

[83] MAGIC Collaboration et al. "Bounds on Lorentz invariance violation from MAGIC observation of GRB 190114C". In: *Phys. Rev. Lett.* (Jan. 2020), p. 17. arXiv: 2001.09728. URL: http://arxiv.org/abs/2001.09728.

[84] Pankaj Mehta et al. *A high-bias, low-variance introduction to Machine Learning for physicists*. 2019. DOI: 10.1016/j.physrep.2019.03.001.

[85] P. Meszaros and M. J. Rees. "Optical and Long-Wavelength Afterglow from Gamma-Ray Bursts". In: *Astrophys. J.* 476.1 (Feb. 1997), pp. 232–237. ISSN: 0004-637X. DOI: 10.1086/303625. arXiv: 9606043 [astro-ph].

[86] Razmik Mirzoyan. *ATel #12390: First time detection of a GRB at sub-TeV energies; MAGIC detects the GRB 190114C*. URL: http://www.astronomerstelegram.org/?read=12390 (visited on 01/30/2020).

[87] Razmik Mirzoyan and Eckart Lorenz. "On the calibration accuracy of light sensors in atmospheric Cherenkov fluorescence and neutrino experiment". In: *Proc. 25th Int. Cosm. Ray Conf.* 7 (1997), pp. 265–268.

[88] D. Morcuende et al. "Relevance of the fluorescence radiation in VHE gamma-ray observations with the Cherenkov technique". In: *Astropart. Phys.* (2019). ISSN: 09276505. DOI: 10.1016/j.astropartphys.2018.11.003.

[89] Mathieu de Naurois and Daniel Mazin. *Ground-based detectors in very-high-energy gamma-ray astronomy*. Aug. 2015. DOI: 10.1016/j.crhy.2015.08.011. arXiv: 1511.00463.

[90] Jun Nishimura and Koichi Kamata. *On the Theory of Cascade Showers, I*. Tech. rep. 2. Feb. 1952, pp. 185–192. DOI: 10.1143/PTP/7.2.185. URL: https://academic.oup.com/ptp/article-abstract/7/2/185/1853403.

[91] Randal S. Olson et al. "Data-driven Advice for Applying Machine Learning to Bioinformatics Problems". In: (2017). arXiv: 1708.05070.

[92] Bohdan Paczynski and James E. Rhoads. "Radio Transients from Gamma-Ray Bursters". In: *Astrophys. J.* 418 (Nov. 1993), p. L5. ISSN: 0004-637X. DOI: 10.1086/187102.

[93] Carsten Peterson, Thorsteinn Rögnvaldsson, and Leif Lönnblad. "JETNET 3.0-A versatile artificial neural network package". In: *Comput. Phys. Commun.* 81.1-2 (June 1994), pp. 185–220. ISSN: 00104655. DOI: 10.1016/0010-4655(94)90120-1.

[94] Viatcheslav I. Priimenko et al. *Inverse Problems of Mathematical Physics*. DE GRUYTER, Mar. 2012. DOI: 10.1515/9783110915525.

[95] Bruno Rossi and Kenneth Greisen. *Cosmic-ray theory*. 1941. DOI: 10.1103/RevModPhys.13.240.

[96] M. Ruderman. "THEORIES OF GAMMA-RAY BURSTS". In: *Ann. N. Y. Acad. Sci.* 262.1 Seventh Texas (Oct. 1975), pp. 164–180. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.1975.tb31430.x. URL: http://doi.wiley.com/10.1111/j.1749-6632.1975.tb31430.x.

[97] George B. Rybicki and Alan P. Lightman. *Radiative Processes in Astrophysics*. Vol. 31. 10. 1979, pp. 359–359. ISBN: 9780471827597. DOI: 10.1088/0031-9112/31/10/053.

[98] Re'em Sari. "The Observed Size and Shape of Gamma-Ray Burst Afterglow". In: *Astrophys. J.* 494.1 (Feb. 1998), pp. L49–L52. ISSN: 0004637X. DOI: 10.1086/311160.

[99] Re'em Sari and Ann A. Esin. "On the Synchrotron Self-Compton Emission from Relativistic Shocks and Its Implications for Gamma-Ray Burst Afterglows". In: *Astrophys. J.* 548.2 (Feb. 2001), pp. 787–799. ISSN: 0004-637X. DOI: 10.1086/319003.

[100] Michael Schmelling. "The method of reduced cross-entropy A general approach to unfold probability distributions". In: *Nucl. Inst. Methods Phys. Res. A* 340.2 (Feb. 1994), pp. 400–412. ISSN: 01689002. DOI: 10.1016/0168-9002(94)90119-8.

[101] J. Selsing. "GRB 190114C: NOT optical counterpart and redshift". In: *GCN Circ.* gcn3 (2019.23695 (2019).

[102] Pierre Sokolsky. "Introduction to ultrahigh energy cosmic ray physics." In: *Front. Physics, Vol. 76,* 76 (1989).

[103] M. Sparre et al. "Spectroscopic evidence for SN 2010ma associated with GRB 101219B". In: *Astrophys. J. Lett.* 735.1 (July 2011), p. L24. ISSN: 20418205. DOI: 10.1088/2041-8205/735/1/L24.

[104] Todor Stanev. *High energy cosmic rays*. Springer, 2010, p. 333. ISBN: 9783540851479.

[105] Ronald C. Stone. "An Accurate Method for Computing Atmospheric Refraction". In: *Publ. Astron. Soc. Pacific* 108 (Nov. 1996), p. 1051. ISSN: 0004-6280. DOI: 10.1086/133831.

[106] M. Tanabashi et al. *Review of Particle Physics*. Aug. 2018. DOI: 10.1103/PhysRevD.98.030001.

[107] M. Tavani et al. "Discovery of powerful gamma-ray flares from the Crab Nebula". In: *Science (80-.).* 331.6018 (Feb. 2011), pp. 736–739. ISSN: 00368075. DOI: 10.1126/science.1200083.

[108] Diego Tescaro. "TeV gamma-ray observations of nearby Active Galactic Nuclei with the MAGIC telescope: exploring the high energy region of the multiwavelength picture". PhD thesis. 2010.

[109] The Fermi-LAT Collaboration. "Search for Spectral Irregularities due to Photon-Axionlike-Particle Oscillations With the Fermi Large Area Telescope". In: *Phys. Rev. Lett.* 116.16 (Mar. 2016), pp. 1–18. ISSN: 10797114. DOI: 10.1103/PhysRevLett.116.161101. arXiv: 1603.06978. URL: http://arxiv.org/abs/1603.06978%20http://dx.doi.org/10.1103/PhysRevLett.116.161101.

[110] Virginia Trimble. "The Distance to the Crab Nebula and NP 0532". In: *Publ. Astron. Soc. Pacific* 85.507 (Oct. 1973), p. 579. ISSN: 0004-6280. DOI: 10.1086/129507.

[111] P. Veres et al. "Observation of inverse Compton emission from a long γ-ray burst". In: *Nature* 575.7783 (2019), pp. 459–463. ISSN: 0028-0836. DOI: 10.1038/s41586-019-1754-6. URL: http://www.nature.com/articles/s41586-019-1754-6.

[112] V. A. Villar et al. "The Combined Ultraviolet, Optical, and Near-infrared Light Curves of the Kilonova Associated with the Binary Neutron Star Merger GW170817: Unified Data Set, Analytic Models, and Physical Implications". In: *Astrophys. J.* 851.1 (Dec. 2017), p. L21. ISSN: 2041-8213. DOI: 10.3847/2041-8213/aa9c84. arXiv: 1710.11576.

[113] Robert Marcus Wagner. "Measurement of Very High Energy Gamma-Ray Emission from Four Blazars Using the MAGIC Telescope and a Comparative Blazar Study". PhD thesis. 2008. DOI: 10.1086/522380. URL: https://publications.mppmu.mpg.de/2006/MPP-2006-245/FullText.pdf.

[114] T. C. Weekes et al. "Observation of TeV gamma rays from the Crab nebula using the atmospheric Cerenkov imaging technique". In: *Astrophys. J.* 342 (July 1989), p. 379. ISSN: 0004-637X. DOI: 10.1086/167599.

[115] Denis Wouters and Pierre Brun. "Irregularity in gamma ray source spectra as a signature of axionlike particles". In: *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* 86.4 (2012). ISSN: 15507998. DOI: 10.1103/PhysRevD.86.043005. arXiv: 1305.4098.

[116] Roberta Zanin et al. "MARS, the MAGIC analysis and reconstruction software". In: *Proc 33rd ICRC, Rio Janeiro, Brazil, Id. 773, 2013* (2013). URL: http://inspirehep.net/record/1412925/files/icrc2013-0773.pdf.

[117] Bing Zhang. *The Physics of Gamma-Ray Bursts*. Cambridge University Press, Dec. 2018. DOI: 10.1017/9781139226530.

[118] Bing Zhang and Peter Meszaros. "High-Energy Spectral Components in Gamma-Ray Burst Afterglows". In: *Astrophys. J.* 559.1 (Sept. 2001), pp. 110–122. ISSN: 0004-637X. DOI: 10.1086/322400.

[119] Bing Zhang et al. " Physical Processes Shaping Gamma-Ray Burst X-Ray Afterglow Light Curves: Theoretical Implications from the Swift X-Ray Telescope Observations ". In: *Astrophys. J.* 642.1 (May 2006), pp. 354–370. ISSN: 0004-637X. DOI: 10.1086/500723. arXiv: 0508321 [astro-ph].

[120] Haozhe Zhang et al. "Random Forest Prediction Intervals". In: *Am. Stat.* (June 2019), pp. 1–15. ISSN: 15372731. DOI: 10.1080/00031305.2019.1585288. URL: https://www.tandfonline.com/doi/full/10.1080/00031305.2019.1585288.

Acknowledgements

I would like to thank all the people who helped me. Here I try to list them, I am afraid

First of all, I would like to thank Masahiro Teshima. He provided me with the opportunity to join the high energy gamma-ray group in ICRR, then to the Max-Planck-Institut für Physik to conduct this research in the MAGIC collaboration. In this research, his response to my opinion was one of the most important guide to look for a next direction in a heavy fog. The logarithmic target was born through the discussion with him.

I am also deeply grateful to Thomas Kuhr. He kindly accepted the supervision of my research despite the different field. And every time when I reported the status, the discussions were very constructive that I could see the missing parts in my research. He also gave me precious ideas. The nested RF was born through the discussion with him.

The daily supervision was given by David Paneque. I owe a great deal to him, especially for supporting the discussions in the collaboration to make the RF-Erec standard, and for revising this thesis into a good shape. I was also helped by him very much to find a peculiarity which needs further investigation. Many trials of getting questions from him and finding an answer clarified various kinds of mysteries like many systematic effects and the effect of RF-Erec on the GRB spectral analysis.

I was with assistance by Chai Yating. She helped me with the MC simulations for understanding the shower development and the Cherenkov light propagation. I could deepen my knowledge about the EAS and the response of IACTs, not only for γ-rays, but also for hadronic showers and muons. It was a great help that the simulation also covered the high Zd observation conditions.

Abelardo Moralejo helped me with deepening my insights on the systematic uncertainties. The relation between the population of the train data and the estimation. His question, "what happens if the target is Energy/Size?" was one example.

As far as I remember, Koji Noda was the only person who believed the potential of the RF-Erec before I got the first good result. I was helped by his encouragement at that moment.

Darko Zarić was the first person who brought me the visible systematic effect in very high Zd observation. Since then my broader search for systematic effects had started.

Martin Will and Alexander Hahn also gave me extraordinary efforts for checking the manuscript of this thesis. Most of the knowledge of the details about the MAGIC telescopes and the observations are given also by them, through the observation shifts.

I thank Diana Werner, the secretary in this group in MPI, for dealing with many issues of bureaucracy, especially for the extended research term.

I was supported by Giovanni Ceribella, especially for mathematical discussions, e.g. theoretical background of the error bar of RMS.

Mathematical support was given also by Ievgen Vovk, e.g., the error bar of a residual value after a subtraction of two weighted values. I got also a sense of pair plots from him.

Petar Temnikov kindly provided me with the ANN program for MAGIC analysis. The comparison of different ML algorithm formed essential part of my understanding of ML, hence to the systematic effects.

What I have done here is simply unbelievable. I had not imagined me conducting, and even completing a doctoral course in physics. This is the consequence of many dramatic changes of the direction of my life. In this regard, I must thank all the opportunities to led me here. But I must thank also all the difficulties and failures, because, although it is embarrassing to say, they were the primary motivations to change the directions of my life at those moments, which finally guided me here.